02-107010US
0186.210

PATENT APPLICATION

NOVEL GLYPHOSATE N-ACETYLTRANSFERASE (GAT) GENES

Inventor(s):

Linda A. Castle, a citizen of the United States
residing at: 784 Hans Avenue, Mountain View, CA 94040

Dan Siehl, a citizen of the United States
residing at: 310 Durham Street, Menlo Park, CA 94025

Lorraine Giver, a citizen of the United States,
residing at: 2538 Hawkington Ct., Santa Clara, CA 95051

Jeremy Minshull, a citizen of the United Kingdom,
residing at: 842 Hermosa Way, Menlo Park, CA 94025

Christina Ivy, a citizen of the United States
residing at 1349 Flora Avenue, San Jose, CA 95130

Yong Hong Chen, a citizen of China
residing at 1828 Vega Circle, Foster City, CA 94404

Nicholas B. Duck, a citizen of the United States,
residing at 1227 Tartarian Trail, Apex, NC 27502

Assignee: Maxygen, Inc.

515 Galveston Drive Redwood City, CA 94063

Entity: **Large**

THE LAW OFFICES OF JONATHAN ALAN QUINE

P.O. Box 458
Alameda, CA 94501
Internet address: www.quinelaw.com

Phone: (510) 337-7871
Fax: (510) 337-7877
E-mail: jaquine@quinelaw.com

02-107010US

NOVEL GLYPHOSATE N-ACETYLTRANSFERASE (GAT) GENES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to and benefit of U.S. Provisional Patent Application Serial No. 60/244,385 filed October 30, 2000, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

COPYRIGHT NOTIFICATION PURSUANT TO 37 C.F.R. § 1.71(E)

BACKGROUND OF THE INVENTION

Crop selectivity to specific herbicides can be conferred by engineering genes into crops which encode appropriate herbicide metabolizing enzymes. In some cases these enzymes, and the nucleic acids that encode them, originate in a plant. In other cases, they are derived from other organisms, such as microbes. *See*, e.g., Padgette et al. (1996) "New weed control opportunities: Development of soybeans with a Round UP Ready™ gene" in Herbicide-Resistant Crops (Duke, ed.), pp54-84, CRC Press, Boca Raton; and Vasil (1996) "Phosphinothricin-resistant crops" in Herbicide-Resistant Crops (Duke, ed.), pp85-91. Indeed, transgenic plants have been engineered to express a variety of herbicide tolerance/metabolizing genes, from a variety of organisms. For example, acetohydroxy acid synthase, which has been found to make plants that express this enzyme resistant to multiple types of herbicides, has been introduced into a variety of plants (*see*, e.g., Hattori et al. (1995) Mol Gen Genet 246:419. Other genes that confer tolerance to herbicides include: a gene encoding a chimeric protein of rat cytochrome P4507A1 and yeast NADPH-cytochrome P450 oxidoreductase (Shiota et al. (1994) Plant PhysiolPlant Physiol 106:17), genes for glutathione reductase and superoxide dismutase (Aono et al. (1995) Plant Cell Physiol 36:1687, and genes for various phosphotransferases (Datta et al. (1992) Plant Mol Biol 20:619.

One herbicide which is the subject of much investigation in this regard is N-phosphonomethylglycine, commonly referred to as glyphosate. Glyphosate is the top selling herbicide in the world, with sales projected to reach $5 billion by 2003. It is a broad spectrum herbicide that kills both broadleaf and grass-type plants. A successful mode of commercial level glyphosate resistance in transgenic plants is by introduction of a modified Agrobacterium CP4 5-enolpyruvylshikimate-3-phosphate synthase (hereinafter referred to as EPSP synthase or EPSPS) gene. The transgene is targeted to the chloroplast where it is capable of continuing to synthesize EPSP from phosphoenolpyruvic acid (PEP) and shikimate-3-phosphate in the presence of glyphosate. In contrast, the native EPSP synthase is inhibited by glyphosate. Without the transgene, plants sprayed with glyphosate quickly die due to inhibition of EPSP synthase which halts the downstream pathway needed for aromatic amino acid, hormone, and vitamin biosynthesis. The CP4 glyphosate-resistant soybean transgenic plants are marketed, e.g., by Monsanto under the name "Round UP Ready™."

In the environment, the predominant mechanism by which glyphosate is degraded is through soil microflora metabolism. The primary metabolite of glyphosate in soil has been identified as aminomethylphosphonic acid (AMPA), which is ultimately converted into ammonia, phosphate and carbon dioxide. The proposed metabolic scheme that describes the degradation of glyphosate in soil through the AMPA pathway is shown in Fig. 8. An alternative metabolic pathway for the breakdown of glyphosate by certain soil bacteria, the sarcosine pathway, occurs via initial cleavage of the C-P bond to give inorganic phosphate and sarcosine, as depicted in Fig. 9.

Another successful herbicide/transgenic crop package is glufosinate (phosphinothricin) and the LibertyLink™ trait marketed, e.g., by Aventis. Glufosinate is also a broad spectrum herbicide. Its target is the glutamate synthase enzyme of the chloroplast. Resistant plants carry the bar gene from Streptomyces hygroscopicus and achieve resistance by the N-acetylation activity of bar, which modifies and detoxifies glufosinate.

An enzyme capable of acetylating the primary amine of AMPA is reported in PCT Application No. WO00/29596. The enzyme was not described as being able to acetylate a compound with a secondary amine (e.g., glyphosate).

While a variety of herbicide resistance strategies are available as noted above, aditional approaches would have considerable commercial value. The present

invention provides, e.g., novel polynucleotides and polypeptides for conferring herbicide tolerance, as well as numerous other benefits as will become apparent during review of the disclosure.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide methods and reagents for rendering an organism, such as a plant, resistant to glyphosate. This and other objects of the invention are provided by one or more of the embodiments described below.

One embodiment of the invention provides novel polypeptides referred to herein as GAT polypeptides. GAT polypeptides are characterized by their structural similarity to one another, e.g., in terms of sequence similarity when the GAT polypeptides are aligned with one another. Some GAT polypeptides possess glyphosate N-acetyl transferase activity, i.e., the ability to catalyze the acetylation of glyphosate. Some GAT polypeptides are also capable of catalyzing the acetylation of glyphosate analogs and or glyphosate metabolites, e.g., aminomethylphosphonic acid.

Also provided are novel polynucleotides referred to herein as GAT polynucleotides. GAT polynucleotides are characterized by their ability to encode GAT polypeptides. In some embodiments of the invention, a GAT polynucleotide is engineered for better plant expression by replacing one or more parental codons with a synonymous codon that is preferentially used in plants relative to the parental codon. In other embodiments, a GAT polynucleotide is modified by the introduction of a nucleotide sequence encoding an N-terminal chloroplast transit peptide.

GAT polypeptides, GAT polynucleotides and glyphosate N-acetyl transferase activity are described in more detail below. The invention further includes certain fragments of the GAT polypeptides and GAT polynucleotides described herein.

The invention includes non-native variants of the polypeptides and polynucleotides described herein, wherein one or more amino acids of the encoded polypeptide have been mutated.

The invention further provides a nucleic acid construct comprising a polynucleotide of the invention. The construct can be a vector, such as a plant transformation vector. In some aspects a vector of the invention will comprise a T-DNA sequence. The construct can optionally include a regulatory sequence (e.g., a promoter) operably linked to a GAT polynucleotide, where the promoter is heterologous with respect to the polynucleotide and effective to cause sufficient expression of the encoded

polypeptide to enhance the glyphosate tolerance of a plant cell transformed with the nucleic acid construct.

In some aspects of the invention, a GAT polynucleotide functions as a selectable marker, e.g., in a plant, bacteria, actinomycetes, yeast, algae or other fungi. For example, an organism that has been transformed with a vector including a GAT polynucleotide selectable marker can be selected based on its ability to grow in the presence of glyphosate. A GAT marker gene can be used for selection or screening for transformed cells expressing the gene.

The invention further provides vectors with stacked traits, i.e., vectors that encode a GAT and that also include a second polynucleotide sequence encoding a second polypeptide that confers a detectable phenotypic trait upon a cell or organism expressing the second polypeptide at an effective level. The detectable phenotypic trait can function as a selectable marker, e.g, by conferring herbicide resistance, pest resistance, or providing some sort of visible marker.

In one embodiment, the invention provides a composition comprising two or more polynucleotides of the invention.

Compositions containing two or more GAT polynucleotides or encoded polypeptides are a feature of the invention. In some cases, these compositions are libraries of nucleic acids containing, e.g., at least 3 or more such nucleic acids. Compositions produced by digesting the nucleic acids of the invention with a restriction endonuclease, a DNAse or an RNAse, or otherwise fragmenting the nucleic acids, e.g., mechanical shearing, chemical cleavage, etc., are also a feature of the invention, as are compositions produced by incubating a nucleic acid of the invention with deoxyribonucleotide triphosphates and a nucleic acid polymerase, such as a thermostable nucleic acid polymerase.

Cells transduced by a vector of the invention, or which otherwise incorporate the nucleic acid of the invention, are an aspect of the invention. In a preferred embodiment, the cells express a polypeptide encoded by the nucleic acid.

In some embodiments, the cells incorporating the nucleic acids of the invention are plant cells. Transgenic plants, transgenic plant cells and transgenic plant explants incorporating the nucleic acids of the invention are also a feature of the invention. In some embodiments, the transgenic plants, trangenic plant cells or transgenic plant explants express an exogenous polypeptide with glyphosate N-acetyltransferase activity

encoded by the nucleic acid of the invention. The invention also provides transgenic seeds produced by the transgenic plants of the invention.

The invention further provides transgenic plants or transgenic plant explants having enhanced tolerance to glyphosate due to the expression of a polypeptide with glyphosate N-acetyltransferase activity and a polypeptide that imparts glyphosate tolerance by another mechanism, such as, a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase and/or a glyphosate-tolerant glyphosate oxido-reductase. In a further embodiment, the invention provides transgenic plants or transgenic plant explants having enhanced tolerance to glyphosate, as well as tolerance to an additional herbicide due to the expression of a polypeptide with glyphosate N-acetyltransferase activity, a polypeptide that imparts glyphosate tolerance by another mechanism, such as, a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase and/or a glyphosate-tolerant glyphosate oxido-reductase and a polypeptide imparting tolerance to the additional herbicide, such as, a mutated hydroxyphenylpyruvatedioxygenase, a sulfonamide-tolerant acetolactate synthase, a sulfonamide-tolerant acetohydroxy acid synthase, an imidazolinone-tolerant acetolactate synthase, an imidazolinone-tolerant acetohydroxy acid synthase, a phosphinothricin acetyl transferase and a mutated protoporphyrinogen oxidase.

The invention also provides transgenic plants or transgenic plant explants having enhanced tolerance to glyphosate, as well as tolerance to an additional herbicide due to the expression of a polypeptide with glyphosate N-acetyltransferase activity and a polypeptide imparting tolerance to the additional herbicide, such as, a mutated hydroxyphenylpyruvatedioxygenase, a sulfonamide-tolerant acetolactate synthase, a sulfonamide-tolerant acetohydroxy acid synthase, an imidazolinone-tolerant acetolactate synthase, an imidazolinone-tolerant acetohydroxy acid synthase, a phosphinothricin acetyl transferase and a mutated protoporphyrinogen oxidase.

Methods of producing the polypeptides of the invention by introducing the nucleic acids encoding them into cells and then expressing and recovering them from the cells or culture medium are a feature of the invention. In preferred embodiments, the cells expressing the polypeptides of the invention are transgenic plant cells.

Polypeptides that are specifically bound by a polyclonal antisera that reacts against an antigen derived from SEQ ID NOS:6-10 and 263-514, but not to a naturally occuring related sequence, e.g., such as a peptide represented by a subsequence of GenBank accession number CAA70664, as well as antibodies which are produced by administering an antigen derived from any one or more of SEQ ID NOS:6-10 and 263-514

and/or which bind specifically to such antigens and which do not specifically bind to a naturally occuring polypeptide corresponding to GenBank accession number CAA70664, are all features of the invention.

Another aspect of the invention relates to methods of polynucleotide diversification to produce novel GAT polynucleotides and polypeptides by recombining or mutating the nucleic acids of the invention in vitro or in vivo. In an embodiment, the recombination produces at least one library of recombinant GAT polynucleotides. The libraries so produced are embodiments of the invention, as are cells comprising the libraries. Furthermore, methods of producing a modified GAT polynucleotide by mutating a nucleic acid of the invention are embodiments of the invention. Recombinant and mutant GAT polynucleotides and polypeptides produced by the methods of the invention are also embodiments of the invention.

In some aspects of the invention, diversification is achieved by using recursive recombination, which can be accomplished in vitro, in vivo, in silico, or a combination thereof. Some examples of diversification methods described in more detail below are family shuffling methods and synthetic shuffling methods.

The invention provides methods for producing a glyphosate resistant transgenic plant or plant cell that involve transforming a plant or plant cell with a polynucleotide encoding a glyphosate N-acetyltransferase, and optionally regenerating a transgenic plant from the transformed plant cell. In some aspects the polynucleotide is a GAT polynucleotide, optionally a GAT polynucleotide derived from a bacterial source. In some aspects of the invention, the method can comprise growing the transformed plant or plant cell in a concentration of glyphosate that inhibits the growth of a wild-type plant of the same species without inhibiting the growth of the transformed plant. The method can comprise growing the transformed plant or plant cell or progeny of the plant or plant cell in increasing concentrations of glyphosate and/or in a concentration of glyphosate that is lethal to a wild-type plant or plant cell of the same species.

A glyphosate resistant transgenic plant produced by this method can be propagated, for example by crossing it with a second plant, such that at least some progeny of the cross display glyphosate tolerance.

The invention further provides methods for selectively controlling weeds in a field containing a crop that involve planting the field with crop seeds or plants which are glyphosate-tolerant as a result of being transformed with a gene encoding a glyphosate N-

acteyltransferase, and applying to the crop and weeds in the field a sufficient amount of glyphosate to control the weeds without significantly affecting the crop.

The invention further provides methods for controlling weeds in a field and preventing the emergence of glyphosate resistant weeds in a field containing a crop which involve planting the field with crop seeds or plants that are glyphosate tolerant as a result of being transformed with a gene encoding a glyphosate N-acetyltransferase and a gene encoding a polypeptide imparting glyphosate tolerance by another mechanism, such as, a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase and/or a glyphosate-tolerant glyphosate oxido-reductase and applying to the crop and the weeds in the field a sufficient amount of glyphosate to control the weeds without significantly affecting the crop.

In a further embodiment the invention provides methods for controlling weeds in a field and preventing the emergence of herbicide resistant weeds in a field containing a crop which involve planting the field with crop seeds or plants that are glyphosate tolerant as a result of being transformed with a gene encoding a glyphosate N-acetyltransferase, a gene encoding a polypeptide imparting glyphosate tolerance by another mechanism, such as, a glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthase and/or a glyphosate-tolerant glyphosate oxido-reductase and a gene encoding a polypeptide imparting tolerance to an additional herbicide, such as, a mutated hydroxyphenylpyruvatedioxygenase, a sulfonamide-tolerant acetolactate synthase, a sulfonamide-tolerant acetohydroxy acid synthase, an imidazolinone-tolerant acetolactate synthase, an imidazolinone-tolerant acetohydroxy acid synthase, a phosphinothricin acetyl transferase and a mutated protoporphyrinogen oxidase and applying to the crop and the weeds in the field a sufficient amount of glyphosate and an additional herbicide, such as, a hydroxyphenylpyruvatedioxygenase inhibitor, sulfonamide, imidazolinone, bialaphos, phosphinothricin, azafenidin, butafenacil, sulfosate, glufosinate, and a protox inhibitor to control the weeds without significantly affecting the crop.

The invention further provides methods for controlling weeds in a field and preventing the emergence of herbicide resistant weeds in a field containing a crop which involve planting the field with crop seeds or plants that are glyphosate tolerant as a result of being transformed with a gene encoding a glyphosate N-acetyltransferase and a gene encoding a polypeptide imparting tolerance to an additional herbicide, such as, a mutated hydroxyphenylpyruvatedioxygenase, a sulfonamide-tolerant acetolactate synthase, a sulfonamide-tolerant acetohydroxy acid synthase, an imidazolinone-tolerant acetolactate

synthase, an imidazolinone-tolerant acetohydroxy acid synthase, a phosphinothricin acetyl transferase and a mutated protoporphyrinogen oxidase and applying to the crop and the weeds in the field a sufficient amount of glyphosate and an additional herbicide, such as, a hydroxyphenylpyruvatedioxygenase inhibitor, sulfonamide, imidazolinone, bialaphos, phosphinothricin, azafenidin, butafenacil, sulfosate, glufosinate, and a protox inhibitor to control the weeds without significantly affecting the crop.

The invention further provides methods for producing a genetically transformed plant that is tolerant toward glyphosate that involve inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising: (i) a promoter which functions in plant cells to cause the production of an RNA sequence;(ii) a structural DNA sequence that causes the production of an RNA sequence which encodes a GAT; and (iii) a 3' non-translated region which functions in plant cells to cause the addition of a stretch of polyadenyl nucleotides to the 3' end of the RNA sequence;
where the promoter is heterologous with respect to the structural DNA sequence and adapted to cause sufficient expression of the encoded polypeptide to enhance the glyphosate tolerance of a plant cell transformed with the DNA molecule; obtaining a transformed plant cell; and regenerating from the transformed plant cell a genetically transformed plant which has increased tolerance to glyphosate.

The invention further provides methods for producing a crop that involve growing a crop plant that is glyphosate-tolerant as a result of being transformed with a gene encoding a glyphosate N-acteyltransferase, under conditions such that the crop plant produces a crop; and harvesting a crop from the crop plant. These methods often include applying glyphosate to the crop plant at a concentration effective to control weeds. Exemplary crop plants include cotton, corn, and soybean.

The invention also provides computers, computer readable medium and integrated systems, including databases that are composed of sequence records including character strings corresponding to SEQ ID NOs:1-514. Such integrated systems optionally include, one or more instruction set for selecting, aligning, translating,reverse-translating or viewing any one or more character strings corresponding to SEQ ID NOs:1-514, with each other and/or with any additional nucleic acid or amino acid sequence.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 depicts the N-acetylation of glyphosate catalyzed by a glyphosate-N-acetyltransferase ("GAT").

Figure 2 illustrates mass spectroscopic detection of N-acetylglyphosate produced by an exemplary Bacillus culture expressing a native GAT activity.

Figure 3 is a table illustrating the relative identity between GAT sequences isolated from different strains of bacteria and yitI from Bacillus subtilis.

Figure 4 is a map of the plasmid pMAXY2120 for expression and purification of the GAT enzyme from E. coli cultures.

Figure 5 is a mass spectrometry output showing increased N-acetylglyphosate production over time in a typical GAT enzyme reaction mix.

Figure 6 is a plot of the kinetic data of a GAT enzyme from which a K_M of 2.9 mM for glyphosate was calculated.

Figure 7 is a plot of the kinetic data taken from the data of Figure 6 from which a K_M of 2 μM was calculated for Acetyl CoA.

Figure 8 is a scheme that describes the degradation of glyphosate in soil through the AMPA pathway.

Figure 9 is a scheme that describes the sarcosine pathway of glyphosate degradation.

Figure 10 is the BLOSUM62 matrix.

Figure 11 is a map of the plasmid pMAXY2190.

Figure 12 depicts a T-DNA construct with *gat* selectable marker.

Figure 13 depicts a yeast expression vector with *gat* selectable marker.

DETAILED DISCUSSION

The present invention relates to a novel class of enzymes exhibiting N-acetyltransferase activity. In one aspect, the invention relates to a novel class of enzymes capable of acetylating glyphosate and glyphosate analogs, e.g., enzymes possessing glyphosate N-acetyltransferase ("GAT") activity. Such enzymes are characterized by the ability to acetylate the secondary amine of a compound. In some aspects of the invention, the compound is a herbicide, e.g., glyphosate, as illustrated schematically in Figure 1. The compound can also be a glyphosate analog or a metabolic product of glyphosate degradation, e.g, aminomethylphosphonic acid. Although the acetylation of glyphosate is a key catalytic step in one metabolic pathway for catabolism of glyphosate, the enzymatic acetylation of glyphosate by naturally-occurring, isolated, or recombinant enzymes has not been previously described. Thus, the nucleic acids and polypeptides of the invention provide a new biochemical pathway for engineering herbicide resistance.

In one aspect, the invention provides novel genes encoding GAT polypeptides. Isolated and recombinant GAT polynucleotides corresponding to naturally occurring polynucleotides, as well as recombinant and engineered, e.g., diversified, GAT polynucleotides are a feature of the invention. GAT polynucleotides are exemplified by SEQ ID NOS: 1-5 and 11-262. Specific GAT polynucleotide and polypeptide sequences are provided as examples to help illustrate the invention, and are not intended to limit the scope of the genus of GAT polynucleotides and polypeptides described and/or claimed herein.

The invention also provides methods for generating and selecting diversified libraries to produce additional GAT polynucleotides, including polynucleotides encoding GAT polypeptides with improved and/or enhanced characteristics, e.g., altered Km for glyphosate, increased rate of catalysis, increased stability, etc., based upon selection of a polynucleotide constituent of the library for the new or improved activities described herein. Such polynucleotides are especially favorably employed in the production of glyphosate resistant transgenic plants.

The GAT polypeptides of the invention exhibit a novel enzymatic activity. Specifically, the enzymatic acetylation of the synthetic herbicide glyphosate has not been recognized prior to the present invention. Thus, the polypeptides herein described, e.g., as exemplified by SEQ ID NOS: 6-10 and 263-514, define a novel biochemical pathway for the detoxification of glyphosate that is functional in vivo, e.g., in plants.

Accordingly, the nucleic acids and polypeptides of the invention are of significant utility in the generation of glyphosate resistant plants by providing new nucleic acids, polypeptides and biochemical pathways for the engineering of herbicide selectivity in transgenic plants.

DEFINITIONS

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular compositions or biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a device" includes a combination of two or more such devices, reference to "a gene fusion construct" includes mixtures of constructs, and the like.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although any methods and materials similar or equivalent to those described herein can be used in the practice for testing of the present invention, specific examples of appropriate materials and methods are described herein.

In describing and claiming the present invention, the following terminology will be used in accordance with the definitions set out below.

For purposes of the present invention, the term "glyphosate" should be considered to include any herbicidally effective form of N-phosphonomethylglycine (including any salt thereof) and other forms which result in the production of the glyphosate anion in planta. The term "glyphosate analog" refers to any structural analog of glyphostate that has the ability to inhibit EPSPS at levels such that the glyphosate analog is herbicidally effective.

As used herein, the term "glyphosate-N-acetyltransferase activity" or "GAT activity" refers to the ability to catalyze the acetylation of the secondary amine group of glyphosate, as illustrated, for example, in Figure 1. A "glyphosate –N-acetyltransferase" or "GAT" is an enzyme that catalyzes the acetylation of the amine group of glyphosate, a glyphosate analog, and/or a glyphosate primary metabolite (i.e., AMPA or sarcosine). In some preferred embodiments of the invention, a GAT is able to transfer the acetyl group from AcetylCoA to the secondary amine of glyphosate and the primary amine of AMPA. The exemplary GATs described herein are active from pH 5-9, with optimal activity in the range of pH 6.5-8.0. Activity can be quantified using various kinetic parameters well know in the art, e.g., k_{cat}, K_M, and k_{cat}/ K_M. These kinetic parameters can be determined as described below in Example 7.

The terms "polynucleotide," "nucleotide sequence," and "nucleic acid" are used to refer to a polymer of nucleotides (A,C,T,U,G, etc. or naturally occurring or artificial nucleotide analogues), e.g., DNA or RNA, or a representation thereof, e.g., a character string, etc, depending on the relevant context. A given polynucleotide or complementary polynucleotide can be determined from any specified nucleotide sequence.

Similarly, an "amino acid sequence" is a polymer of amino acids (a protein, polypeptide, etc.) or a character string representing an amino acid polymer, depending on context. The terms "protein," "polypeptide," and "peptide" are used interchangeably herein.

A polynucleotide, polypeptide or other component is "isolated" when it is partially or completely separated from components with which it is normally associated (other proteins, nucleic acids, cells, synthetic reagents, etc.). A nucleic acid or polypeptide is "recombinant" when it is artificial or engineered, or derived from an artificial or engineered protein or nucleic acid. For example, a polynucleotide that is inserted into a vector or any other heterologous location, e.g, in a genome of a recombinant organism, such that it is not associated with nucleotide sequences that normally flank the polynucleotide as it is found in nature is a recombinant polynucleotide. A protein expressed in vitro or in vivo from a recombinant polynucleotide is an example of a recombinant polypeptide. Likewise, a polynucleotide sequence that does not appear in nature, for example a variant of a naturally occurring gene, is recombinant.

The terms "glyphosate N-acetyl transferase polypeptide" and "GAT polypeptide" are used interchangeably to refer to any of a family of novel polypeptides provided herein.

The terms "glyphosate N-acetyl transferase polynucleotide" and "GAT polynucleotide" are used interchangeably to refer to a polynucleotide that encodes a GAT polypeptide.

A "subsequence" or "fragment" is any portion of an entire sequence.

Numbering of an amino acid or nucleotide polymer corresponds to numbering of a selected amino acid polymer or nucleic acid when the position of a given monomer component (amino acid residue, incorporated nucleotide, etc.) of the polymer corresponds to the same residue position in a selected reference polypeptide or polynucleotide.

A vector is a composition for facilitating cell transduction by a selected nucleic acid, or expression of the nucleic acid in the cell. Vectors include, e.g., plasmids, cosmids, viruses, YACs, bacteria, poly-lysine, chromosome integration vectors, episomal vectors, etc.

"Substantially an entire length of a polynucleotide or amino acid sequence" refers to at least about 70%, generally at least about 80%, or typically about 90% or more of a sequence.

As used herein, an "antibody" refers to a protein comprising one or more polypeptides substantially or partially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as myriad

immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively. A typical immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (VL) and variable heavy chain (VH) refer to these light and heavy chains respectively. Antibodies exist as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab)'2, a dimer of Fab which itself is a light chain joined to VH-CH1 by a disulfide bond. The F(ab)'2 may be reduced under mild conditions to break the disulfide linkage in the hinge region thereby converting the (Fab')2 dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (see, Fundamental Immunology, 4th Edition,W.E. Paul (ed.), Raven Press, N.Y. (1998), for a more detailed description of other antibody fragments). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such Fab' fragments may be synthesized de novo either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein also includes antibody fragments either produced by the modification of whole antibodies or synthesized de novo using recombinant DNA methodologies. Antibodies include single chain antibodies, including single chain Fv (sFv) antibodies in which a variable heavy and a variable light chain are joined together (directly or through a peptide linker) to form a continuous polypeptide.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide.

A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J. J., (1991) Ann. Rev. Plant Phys. Plant Mol. Biol. 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (supra) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (supra) may be added. If the protein

is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel, N. (1992) Plant Phys. 100:1627-1632).

The terms "diversification" and "diversity," as applied to a polynucleotide, refers to generation of a plurality of modified forms of a parental polynucleotide, or plurality of parental polynucleotides. In the case where the polynucleotide encodes a polypeptide, diversity in the nucleotide sequence of the polynucleotide can result in diversity in the corresponding encoded polypeptide, e.g. a diverse pool of polynucleotides encoding a plurality of polypeptide variants. In some embodiments of the invention, this sequence diversity is exploited by screening/selecting a library of diversified polynucleotides for variants with desirable functional attributes, e.g., a polynucleotide encoding a GAT polypeptide with enhanced functional characteristics.

The term "encoding" refers to the ability of a nucleotide sequence to code for one or more amino acids. The term does not require a start or stop codon. An amino acid sequence can be encoded in any one of six different reading frames provided by a polynucleotide sequence and its complement.

When used herein, the term "artificial variant" refers to a polypeptide having GAT activity, which is encoded by a modified GAT polynucleotide, e.g., a modified form of any one of SEQ ID NOS: 1-5 and 11-262, or of a naturally-occurring GAT polynucleotide isolated from an organism. The modified polynucleotide, from which an artificial variant is produced when expressed in a suitable host, is obtained through human intervention by modification of a GAT polynucleotide.

The term "nucleic acid construct" or "polynucleotide construct" means a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature. The term nucleic acid construct is synonymous with the term "expression cassette" when the nucleic acid construct contains the control sequences required for expression of a coding sequence of the present invention.

The term "control sequences" is defined herein to include all components, which are necessary or advantageous for the expression of a polypeptide of the present invention. Each control sequence may be native or foreign to the nucleotide sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and

transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleotide sequence encoding a polypeptide.

The term "operably linked" is defined herein as a configuration in which a control sequence is appropriately placed at a position relative to the coding sequence of the DNA sequence such that the control sequence directs the expression of a polypeptide.

When used herein the term "coding sequence" is intended to cover a nucleotide sequence, which directly specifies the amino acid sequence of its protein product. The boundaries of the coding sequence are generally determined by an open reading frame, which usually begins with the ATG start codon. The coding sequence typically includes a DNA, cDNA, and/or recombinant nucleotide sequence.

In the present context, the term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

In the present context, the term "expression vector" covers a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of the invention, and which is operably linked to additional segments that provide for its transcription.

The term "host cell", as used herein, includes any cell type which is susceptible to transformation with a nucleic acid construct.

The term "plant" includes whole plants, shoot vegetative organs/structures (*e.g.* leaves, stems and tubers), roots, flowers and floral organs/structures (*e.g.* bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (*e.g.* vascular tissue, ground tissue, and the like) and cells (*e.g.* guard cells, egg cells, trichomes and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, and multicellular algae. It includes plants of a variety of ploidy levels, including aneuploid, polyploid, diploid, haploid and hemizygous.

The term "heterologous" as used herein describes a relationship between two or more elements which indicates that the elemennts are not normally found in

proximity to one another in nature. Thus, for example, a polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is not naturally associated with the promoter (*e.g.* a genetically engineered coding sequence or an allele from a different ecotype or variety). An example of a heterologous polypeptide is a polypeptide expressed from a recombinant polynucleotide in a transgenic organism. Heterologous polynucleotides and polypeptides are forms of recombinant molecules.

A variety of additional terms are defined or otherwise characterized herein.

GLYPHOSATE N-ACETYLTRANSFERASES

In one aspect, the invention provides a novel family of isolated or recombinant enzymes referred to herein as "glyphosate N-acetyltransferases," "GATs ," or "GAT enzymes." GATs are enzymes that have GAT activity, preferably sufficient activity to confer some degree of glyphosate tolerance upon a transgenic plant engineered to express the GAT. Some examples of GATs include GAT polypeptides, described in more detail below.

Of course, GAT-mediated glyphosate tolerance is a complex function of GAT activity, GAT expression levels in the transgenic plant, the particular plant, the nature and timing of herbicide application, etc. One of skill in the art can determine without undue experimentation the level of GAT activity required to effect glyphosate tolerance in a particular context.

GAT activity can be characterized using the conventional kinetic parameters k_{cat}, K_M, and k_{cat} / K_M. k_{cat} can be thought of as a measure of the rate of acetylation, particularly at high substrate concentrations, K_M is a measure of the affinity of the GAT for its substrates (e.g., Acetyl CoA and glyphosate), and k_{cat} / K_M is a measure of catalytic efficiency that takes both substrate affinity and catalytic rate into account – this parameter is particularly important in the situation where the concentration of a substrate is at least partially rate limiting. In general, a GAT with a higher k_{cat} or k_{cat} / K_M is a more efficient catalyst than another GAT with lower k_{cat} or k_{cat} / K_M. A GAT with a lower K_M is a more efficient catalyst than another GAT with a higher K_M. Thus, to determine whether one GAT is more effective than another, one can compare kinetic parameters for the two enzymes. The relative importance of k_{cat}, k_{cat} / K_M and K_M will vary depending upon the

context in which the GAT will be expected to function, e.g., the anticipated effective concentration of glyphosate relative to K_M for glyphosate. GAT activity can also be characterized in terms of any of a number of functional characteristics, e.g., stability, susceptibility to inhibition or activation by other molecules, etc.

GLYPHOSATE N-ACETYLTRANSFERASE POLYPEPTIDES

In one aspect, the invention provides a novel family of isolated or recombinant polypeptides referred to herein as "glyphosate N-acetyltransferase polypeptides" or "GAT polypeptides." GAT polypeptides are characterized by their structural similarity to a novel family of GATs. Many but not all GAT polypeptides are GATs. The distinction is that GATs are defined in terms of function, whereas GAT polypeptides are defined in terms of structure. A subset of the GAT polypeptides consists of those GAT polypeptides that have GAT activity, preferably at a level that will function to confer glyphosate resistance upon a transgenic plant expressing the protein at an effective level. Some preferred GAT polypeptides for use in conferring glyphosate tolerance have a k_{cat} of at least 1 min^{-1}, or more preferably at least 10 min^{-1}, 100 min^{-1} or 1000 min^{-1}. Other preferred GAT polypeptides for use in conferring glyphosate tolerance have a K_M no greater than 100 mM, or more preferably no greater than 10 mM, 1 mM, or 0.1 mM. Still other preferred GAT polypeptides for use in conferring glyphosate tolerance have a k_{cat}/ K_M of at least 1 $mM^{-1}min^{-1}$ or more, preferably at least 10 $mM^{-1}min^{-1}$, 100 $mM^{-1}min^{-1}$, 1000 $mM^{-1}min^{-1}$, or 10,000 $mM^{-1}min^{-1}$.

Exemplary GAT polypeptides have been isolated and characterized from a variety of bacterial strains. One example of a monomeric GAT polypeptide that has been isolated and characterized has a molecular radius of approximately 17 kD. An exemplary GAT enzyme isolated from a strain of *B. licheniformis*, SEQ ID NO:7, exhibits a Km for glyphosate of approximately 2.9 mM and a Km for acetyl CoA of approximately 2 μM, with a kcat equal to 6/minute.

The term "GAT polypeptide" refers to any polypeptide comprising an amino acid sequence that can be optimally aligned with an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514 to generate a similarity score of at least 430 using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence that can be optimally aligned with an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514 to generate a similarity score of at least 440, 445, 450, 455, 460, 465, 470, 475, 480, 485,

490, 495, 500, 505, 510, 515, 520, 525, 530, 535, 540, 545, 550, 555, 560, 565, 570, 575, 580, 585, 590, 595, 600, 605, 610, 615, 620, 625, 630, 635, 640, 645, 650, 655, 660, 665, 670, 675, 680, 685, 690, 695, 700, 705, 710, 715, 720, 725, 730, 735, 740, 745, 750, 755, or 760 using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence that can be optimally aligned with SEQ ID NO. 457 to generate a similarity score of at least 430 using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence that can be optimally aligned with SEQ ID NO. 457 to generate a similarity score of at least 440, 445, 450, 455, 460, 465, 470, 475, 480, 485, 490, 495, 500, 505, 510, 515, 520, 525, 530, 535, 540, 545, 550, 555, 560, 565, 570, 575, 580, 585, 590, 595, 600, 605, 610, 615, 620, 625, 630, 635, 640, 645, 650, 655, 660, 665, 670, 675, 680, 685, 690, 695, 700, 705, 710, 715, 720, 725, 730, 735, 740, 745, 750, 755, or 760 using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence that can be optimally aligned with SEQ ID NO. 445 to generate a similarity score of at least 430 using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence that can be optimally aligned with SEQ ID NO. 445 to generate a similarity score of at least 440, 445, 450, 455, 460, 465, 470, 475, 480, 485, 490, 495, 500, 505, 510, 515, 520, 525, 530, 535, 540, 545, 550, 555, 560, 565, 570, 575, 580, 585, 590, 595, 600, 605, 610, 615, 620, 625, 630, 635, 640, 645, 650, 655, 660, 665, 670, 675, 680, 685, 690, 695, 700, 705, 710, 715, 720, 725, 730, 735, 740, 745, 750, 755, or 760 using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence that can be optimally aligned with SEQ ID NO:300 to generate a similarity score of at least 430 using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence that can be optimally aligned with SEQ ID NO: 300 to generate a similarity score of at least 440, 445, 450, 455, 460, 465, 470, 475, 480, 485, 490, 495, 500, 505, 510, 515, 520, 525, 530, 535, 540, 545, 550, 555, 560,

565, 570, 575, 580, 585, 590, 595, 600, 605, 610, 615, 620, 625, 630, 635, 640, 645, 650, 655, 660, 665, 670, 675, 680, 685, 690, 695, 700, 705, 710, 715, 720, 725, 730, 735, 740, 745, 750, 755, or 760 using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1.

Two sequences are "optimally aligned" when they are aligned for similarity scoring using a defined amino acid substitution matrix (e.g., BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acids substitution matrices and their use in quantifying the similarity between two sequences are well-known in the art and described, e.g., in Dayhoff et al. (1978) "A model of evolutionary change in proteins." In "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M.O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, DC and Henikoff et al. (1992) Proc. Natl. Acad. Sci. USA 89:10915-10919. The BLOSUM62 matrix (Fig. 10) is often used as a default scoring substitution matrix in sequence alignment protocols such as Gapped BLAST 2.0. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The alignment is defined by the amino acids positions of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences, so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is facilitated by the use of a computer-implemented alignment algorithm, e.g., gapped BLAST 2.0, described in Altschul et al, (1997) Nucleic Acids Res. 25:3389-3402, and made available to the public at the National Center for Biotechnology Information Website (http://www.ncbi.nlm.nih.gov). Optimal alignments, including multiple alignments, can be prepared using, e.g., PSI-BLAST, available through http://www.ncbi.nlm.nih.gov and described by Altschul et al, (1997) Nucleic Acids Res. 25:3389-3402.

With respect to an amino acid sequence that is optimally aligned with a reference sequence, an amino acid residue "corresponds to" the position in the reference sequence with which the residue is paired in the alignment. The "position" is denoted by a number that sequentially identifies each amino acid in the reference sequence based on its position relative to the N-terminus. For example, in SEQ ID NO:300 position 1 is M, position 2 is I, position 3 is E, etc. When a test sequence is optimally aligned with SEQ ID NO:300, a residue in the test sequence that aligns with the E at position 3 is said to

"correspond to position 3" of SEQ ID NO:300. Owing to deletions, insertion, truncations, fusions, etc., that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence as determined by simply counting from the N-terminal will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where there is a deletion in an aligned test sequence, there will be no amino acid that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to any amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

The term "GAT polypeptide" further refers to any polypeptide comprising an amino acid sequence having at least 40% sequence identity with an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence having at least 40% sequence identity with SEQ ID NO. 457. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with SEQ ID NO. 457.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence having at least 40% sequence identity with SEQ ID NO. 445. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with SEQ ID NO. 445.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence having at least 40% sequence identity with SEQ ID NO. 300. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with SEQ ID NO. 300.

The term "GAT polypeptide" further refers to any polypeptide comprising an amino acid sequence having at least 40% sequence identity with residues 1-96 of an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514. Some aspects of the invention pertain to polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with residues 1-96 of an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514.

One aspect of the invention pertains to a polypeptide comprising an amino acid sequence having at least 40% sequence identity with residues 1-96 of SEQ ID NO. 457. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with residues 1-96 of SEQ ID NO. 457.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence having at least 40% sequence identity with residues 1-96 of SEQ ID NO. 445. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with residues 1-96 of SEQ ID NO. 445.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence having at least 40% sequence identity with residues 1-96 of SEQ ID NO. 300. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with residues 1-96 of SEQ ID NO. 300.

The term "GAT polypeptide" further refers to any polypeptide comprising an amino acid sequence having at least 40% sequence identity with residues 51-146 of an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514. Some aspects of the invention pertain to polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with residues 51-146 of an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514.

One aspect of the invention pertains to a polypeptide comprising an amino acid sequence having at least 40% sequence identity with residues 51-146 of SEQ ID NO: 457. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with residues 51-146 of SEQ ID NO. 457.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence having at least 40% sequence identity with residues 51-146 of SEQ ID NO. 445. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with residues 51-146 of SEQ ID NO. 445.

One aspect of the invention pertains to a GAT polypeptide comprising an amino acid sequence having at least 40% sequence identity with residues 51-146 of SEQ ID NO. 300. Some aspects of the invention pertain to GAT polypeptides comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% sequence identity with residues 51-146 of SEQ ID NO. 300.

As used herein, the term "identity" or "percent identity" when used with respect to a particular pair of aligned amino acid sequences, refers to the percent amino acid sequence identity that is obtained by ClustalW analysis (version W 1.8 available from European Bioinformatics Institute, Cambridge, UK), counting the number of identical matches in the alignment and dividing such number of identical matches by the greater of (i) the length of the aligned sequences, and (ii) 96, and using the following default ClustalW parameters to achieve slow/accurate pairwise alignments – Gap Open Penalty:10; Gap Extension Penalty:0.10; Protein weight matrix:Gonnet series; DNA weight matrix: IUB; Toggle Slow/Fast pairwise alignments = SLOW or FULL Alignment.

In another aspect, the invention provides an isolated or recombinant polypeptide that comprises at least 20, or alternatively, 50, 75, 100, 125 or 140 contiguous amino acids of an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514.

In another aspect, the invention provides an isolated or recombinant polypeptide that comprises at least 20, or alternatively, 50, 100 or 140 contiguous amino acids of SEQ ID NO:457.

In another aspect, the invention provides an isolated or recombinant polypeptide that comprises at least 20, or alternatively, 50, 100 or 140 contiguous amino acids of SEQ ID NO:445.

In another aspect, the invention provides an isolated or recombinant polypeptide that comprises at least 20, or alternatively, 50, 100 or 140 contiguous amino acids of SEQ ID NO:300.

In another aspect, the invention provides a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514.

Some preferred GAT polypeptides of the invention are characterized as follows. When optimally aligned with a reference amino acid sequence selected from the group consisting of SEQ ID NO:6-10 and 263-514, at least 90% of the amino acid residues in the polypeptide that correspond to the following positions conform to the following restrictions: (a) at positions 2, 4, 15, 19, 26, 28, 31, 45, 51, 54, 86, 90, 91, 97, 103, 105, 106, 114, 123, 129, 139, and/or 145 the amino acid residue is B1; and (b) at positions 3, 5, 8, 10, 11, 14, 17, 18, 24, 27, 32, 37, 38, 47, 48, 49, 52, 57, 58, 61, 62, 63, 68, 69, 79, 80, 82, 83, 89, 92, 100, 101, 104, 119, 120, 124, 125, 126, 128, 131, 143, and/or 144 the amino acid residue is B2; wherein B1 is an amino acid selected from the group consisting of A, I, L, M, F, W, Y, and V; and B2 is an amino acid selected from the group consisting of R, N, D, C, Q, E, G, H, K, P, S, and T. When used to specify an amino acid or amino acid residue, the single letter designations A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, and Y have their standard meaning as used in the art and as provided in Table 2 herein.

Some preferred GAT polypeptides of the invention are characterized as follows. When optimally aligned with a reference amino acid sequence selected from the group consisting of SEQ ID NO:6-10 and 263-514, at least 80% of the amino acid residues in the polypeptide that correspond to the following positions conform to the following restrictions: (a) at positions 2, 4, 15, 19, 26, 28, 51, 54, 86, 90, 91, 97, 103, 105, 106, 114, 129, 139, and/or 145 the amino acid residue is Z1; (b) at positions 31 and/or 45 the amino acid residue is Z2; (c) at positions 8 and/or 89 the amino acid residue is Z3; (d) at positions 82, 92, 101 and/or 120 the amino acid residue is Z4; (e) at positions 3, 11, 27 and/or 79 the amino acid residue is Z5; (f) at position 123 the amino acid residue is Z1 or Z2; (g) at positions 12, 33, 35, 39, 53, 59, 112, 132, 135, 140, and/or 146 the amino acid residue is Z1 or Z3; (h) at position 30 the amino acid residue is Z1 or Z4; (i) at position 6 the amino acid residue is Z1 or Z6; (j) at positions 81 and/or 113 the amino acid residue is Z2 or Z3; (k) at positions 138 and/or 142 the amino acid residue is Z2 or Z4; (l) at positions 5, 17, 24, 57, 61, 124 and/or 126 the amino acid residue is Z3 or Z4; (m) at position 104 the amino acid residue is Z3 or Z5; (o) at positions 38, 52, 62 and/or 69 the amino acid residue is Z3 or Z6; (p) at positions 14, 119 and/or 144 the amino acid residue is Z4 or Z5; (q) at position 18 the amino acid residue is Z4 or Z6; (r) at positions 10, 32,

48, 63, 80 and/or 83 the amino acid residue is Z5 or Z6; (s) at position 40 the amino acid residue is Z1, Z2 or Z3; (t) at positions 65 and/or 96 the amino acid residue is Z1, Z3 or Z5; (u) at positions 84 and/or 115 the amino acid residue is Z1, Z3 or Z4; (v) at position 93 the amino acid residue is Z2, Z3 or Z4; (w) at position 130 the amino acid residue is Z2, Z4 or Z6; (x) at positions 47 and/or 58 the amino acid residue is Z3, Z4 or Z6; (y) at positions 49, 68, 100 and/or 143 the amino acid residue is Z3, Z4 or Z5; (z) at position 131 the amino acid residue is Z3, Z5 or Z6; (aa) at positions 125 and/or 128 the amino acid residue is Z4, Z5 or Z6; (ab) at position 67 the amino acid residue is Z1, Z3, Z4 or Z5; (ac) at position 60 the amino acid residue is Z1, Z4, Z5 or Z6; and(ad) at position 37 the amino acid residue is Z3, Z4, Z5 or Z6; wherein Z1 is an amino acid selected from the group consisting of A, I, L, M, and V; Z2 is an amino acid selected from the group consisting of F, W, and Y; Z3 is an amino acid selected from the group consisting of N, Q, S, and T; Z4 is an amino acid selected from the group consisting of R, H, and K; Z5 is an amino acid selected from the group consisting of D and E; and Z6 is an amino acid selected from the group consisting of C, G, and P.

Some preferred GAT polypeptides of the invention are characterized as follows. When optimally aligned with a reference amino acid sequence selected from the group consisting of SEQ ID NO:6-10 and 263-514, at least 90% of the amino acid residues in the polypeptide that correspond to the following positions conform to the following restrictions: (a) at positions 1, 7, 9, 13, 20, 36, 42, 46, 50, 56, 64, 70, 72, 75, 76, 78, 94, 98, 107, 110, 117, 118, 121, and/or 141 the amino acid residue is B1; and (b) at positions 16, 21, 22, 23, 25, 29, 34, 41, 43, 44, 55, 66, 71, 73, 74, 77, 85, 87, 88, 95, 99, 102, 108, 109, 111, 116, 122, 127, 133, 134, 136, and/or 137 the amino acid residue is B2; wherein B1 is an amino acid selected from the group consisting of A, I, L, M, F, W, Y, and V; and B2 is an amino acid selected from the group consisting of R, N, D, C, Q, E, G, H, K, P, S, and T.

Some preferred GAT polypeptides of the invention are characterized as follows. When optimally aligned with a reference amino acid sequence selected from the group consisting of SEQ ID NO:6-10 and 263-514, at least 90% of the amino acid residues in the polypeptide that correspond to the following positions conform to the following restrictions: (a) at positions 1, 7, 9, 20, 36, 42, 50, 64, 72, 75, 76, 78, 94, 98, 110, 121, and/or 141 the amino acid residue is Z1; (b) at positions 13, 46, 56, 70, 107, 117, and/or 118 the amino acid residue is Z2; (c) at positions 23, 55, 71, 77, 88, and/or 109 the amino acid residue is Z3; (d) at positions 16, 21, 41, 73, 85, 99, and/or 111 the amino acid residue is Z4; (e) at positions 34 and/or 95 the amino acid residue is Z5; (f) at position 22,

25, 29, 43, 44, 66, 74, 87, 102, 108, 116, 122, 127, 133, 134, 136, and/or 137 the amino acid residue is Z6; wherein Z1 is an amino acid selected from the group consisting of A, I, L, M, and V; Z2 is an amino acid selected from the group consisting of F, W, and Y; Z3 is an amino acid selected from the group consisting of N, Q, S, and T; Z4 is an amino acid selected from the group consisting of R, H, and K; Z5 is an amino acid selected from the group consisting of D and E; and Z6 is an amino acid selected from the group consisting of C, G, and P.

Some preferred GAT polypeptides of the invention are characterized as follows. When optimally aligned with a reference amino acid sequence selected from the group consisting of SEQ ID NO:6-10 and 263-514, at least 80% of the amino acid residues in the polypeptide that correspond to the following positions conform to the following restrictions: (a) at position 2 the amino acid residue is I or L; (b) at position 3 the amino acid residue is E or D; (c) at position 4 the amino acid residue is V, A or I; (d) at position 5 the amino acid residue is K, R or N; (e) at position 6 the amino acid residue is P or L; (f) at position 8 the amino acid residue is N, S or T; (g) at position 10 the amino acid residue is E or G; (h) at position 11 the amino acid residue is D or E; (i) at position 12 the amino acid residue is T or A; (j) at position 14 the amino acid residue is E or K; (k) at position 15 the amino acid residue is I or L; (l) at position 17 the amino acid residue is H or Q; (m) at position 18 the amino acid residue is R, C or K; (n) at position 19 the amino acid residue is I or V; (o) at position 24 the amino acid residue is Q or R; (p) at position 26 the amino acid residue is L or I; (q) at position 27 the amino acid residue is E or D; (r) at position 28 the amino acid residue is A or V; (s) at position 30 the amino acid residue is K, M or R; (t) at position 31 the amino acid residue is Y or F; (u) at position 32 the amino acid residue is E or G; (v) at position 33 the amino acid residue is T, A or S; (w) at position 35 the amino acid residue is L, S or M; (x) at position 37 the amino acid residue is R, G, E or Q; (y) at position 38 the amino acid residue is G or S; (z) at position 39 the amino acid residue is T, A or S; (aa) at position 40 the amino acid residue is F, L or S; (ab) at position 45 the amino acid residue is Y or F; (ac) at position 47 the amino acid residue is R, Q or G; (ad) at position 48 the amino acid residue is G or D; (ae) at position 49 the amino acid residue is K, R, E or Q; (af) at position 51 the amino acid residue is I or V; (ag) at position 52 the amino acid residue is S, C or G; (ah) at position 53 the amino acid residue is I or T; (ai) at position 54 the amino acid residue is A or V; (aj) at position 57 the amino acid residue is H or N; (ak) at position 58 the amino acid residue is Q, K, N or P; (al) at position 59 the amino acid residue is A or S; (am) at position 60 the amino acid residue is E, K, G, V or

D; (an) at position 61 the amino acid residue is H or Q; (ao) at position 62 the amino acid residue is P, S or T; (ap) at position 63 the amino acid residue is E, G or D; (aq) at position 65 the amino acid residue is E, D, V or Q; (ar) at position 67 the amino acid residue is Q, E, R, L, H or K; (as) at position 68 the amino acid residue is K, R, E, or N; (at) at position 69 the amino acid residue is Q or P; (au) at position 79 the amino acid residue is E or D; (av) at position 80 the amino acid residue is G or E; (aw) at position 81 the amino acid residue is Y, N or F; (ax) at position 82 the amino acid residue is R or H; (ay) at position 83 the amino acid residue is E, G or D; (az) at position 84 the amino acid residue is Q, R or L; (ba) at position 86 the amino acid residue is A or V; (bb) at position 89 the amino acid residue is T or S; (bc) at position 90 the amino acid residue is L or I; (bd) at position 91 the amino acid residue is I or V; (be) at position 92 the amino acid residue is R or K; (bf) at position 93 the amino acid residue is H, Y or Q; (bg) at position 96 the amino acid residue is E, A or Q; (bh) at position 97 the amino acid residue is L or I; (bi) at position 100 the amino acid residue is K, R, N or E; (bj) at position 101 the amino acid residue is K or R; (bk) at position 103 the amino acid residue is A or V; (bl) at position 104 the amino acid residue is D or N; (bm) at position 105 the amino acid residue is L or M; (bn) at position 106 the amino acid residue is L or I; (bo) at position 112 the amino acid residue is T or I; (bp) at position 113 the amino acid residue is S, T or F; (bq) at position 114 the amino acid residue is A or V; (br) at position 115 the amino acid residue is S, R or A; (bs) at position 119 the amino acid residue is K, E or R; (bt) at position 120 the amino acid residue is K or R; (bu) at position 123 the amino acid residue is F or L; (bv) at position 124 the amino acid residue is S or R; (bw) at position 125 the amino acid residue is E, K, G or D; (bx) at position 126 the amino acid residue is Q or H; (by) at position 128 the amino acid residue is E, G or K; (bz) at position 129 the amino acid residue is V, I or A; (ca) at position 130 the amino acid residue is Y, H, F or C; (cb) at position 131 the amino acid residue is D, G, N or E; (cc) at position 132 the amino acid residue is I, T, A, M, V or L; (cd) at position 135 the amino acid residue is V, T, A or I; (ce) at position 138 the amino acid residue is H or Y; (cf) at position 139 the amino acid residue is I or V; (cg) at position 140 the amino acid residue is L or S; (ch) at position 142 the amino acid residue is Y or H; (ci) at position 143 the amino acid residue is K, T or E; (cj) at position 144 the amino acid residue is K, E or R; (ck) at position 145 the amino acid residue is L or I; and (cl) at position 146 the amino acid residue is T or A.

Some preferred GAT polypeptides of the invention are characterized as follows. When optimally aligned with a reference amino acid sequence selected from the

group consisting of SEQ ID NO:6-10 and 263-514, at least 80% of the amino acid residues in the polypeptide that correspond to the following positions conform to the following restrictions: (a) at position 9, 76, 94 and 110 the amino acid residue is A; (b) at position 29 and 108 the amino acid residue is C; (c) at position 34 the amino acid residue is D; (d) at position 95 the amino acid residue is E; (e) at position 56 the amino acid residue is F; (f) at position 43, 44, 66, 74, 87, 102, 116, 122, 127 and 136 the amino acid residue is G; (g) at position 41 the amino acid residue is H; (h) at position 7 the amino acid residue is I; (i) at position 85 the amino acid residue is K; (j) at position 20, 36, 42, 50, 72, 78, 98 and 121 the amino acid residue is L; (k) at position 1, 75 and 141 the amino acid residue is M; (l) at position 23, 64 and 109 the amino acid residue is N; (m) at position 22, 25, 133, 134 and 137 the amino acid residue is P; (n) at position 71 the amino acid residue is Q; (o) at position 16, 21, 73, 99 and 111 the amino acid residue is R; (p) at position 55 and 88 the amino acid residue is S; (q) at position 77 the amino acid residue is T; (r) at position 107 the amino acid residue is W; and (s) at position 13, 46, 70, 117 and 118 the amino acid residue is Y.

Some preferred GAT polypeptides of the invention are characterized as follows. When optimally aligned with a reference amino acid sequence selected from the group consisting of SEQ ID NO:6-10 and 263-514, the amino acid residue in the polypeptide that correspond to position 28 is V or A. Valine at the 28 position generally correlates with reduced K_M, while alanine at that position generally correlates with increased k_{cat}. Other preferred GAT polypeptides are characterized by having I27 (i.e., an I at position 27), M30, S35, R37, S39, G48, K49, N57, Q58, P62, Q65, Q67, K68, E83, S89, A96, E96, R101, T112, A114, K119, K120, E128, V129, D131, T131, V134, R144, I145, or T146, or any combination thereof.

Some preferred GAT polypeptides of the invention comprise an amino acid sequence selected from the group consisting of SEQ ID NOS:6-10 and 263-514.

The invention further provides preferred GAT polypeptides that are characterized by a combination of the foregoing amino acid residue position restrictions.

In addition, the invention provides GAT polynucleotides encoding the preferred GAT polypeptides described above, and complementary nucleotide sequences thereof.

Some aspects of the invention pertain particularly to the subset of any of the above-described categories of GAT polypeptides having GAT activity, as described herein. These GAT polypeptides are preferred, for example, for use as agents for

conferring glyphosate resistance upon a plant. Examples of desired levels of GAT activity are described herein.

In one aspect, the GAT polypeptides comprise an amino acid sequence encoded by a recombinant or isolated form of naturally occurring nucleic acids isolated from a natural source, e.g., a bacterial strain. Wild-type polynucleotides encoding such GAT polypeptides may be specifically screened for by standard techniques known in the art. The polypeptides defined by SEQ ID NO:6 to SEQ ID NO:10, for example, were discovered by expression cloning of sequences from *Bacillus* strains exhibiting GAT activity, as described in more detail below.

The invention also includes isolated or recombinant polypeptides which are encoded by an isolated or recombinant polynucleotide comprising a nucleotide sequence which hybridizes under stringent conditions over substantially the entire length of a nucleotide sequence selected from the group consisting of SEQ ID NOS: 1-5 and 11-262, their complements, and nucleotide sequences encoding an amino acid sequence selected from the group consisting of SEQ ID NOS: 6-10 and 263-514, including their complements.

The invention further includes any polypeptide having GAT activity that is encoded by a fragment of any of the GAT-encoding polynucleotides described herein.

The invention also provides fragments of GAT polypeptides that can be spliced together to form a functional GAT polypeptide. Splicing can be accomplished in vitro or in vivo, and can involve cis or trans (i.e., intramolecular or intermolecular) splicing. The fragments themselves can, but need not, have GAT activity. For example, two or more segments of a GAT polypeptide can be separated by inteins; removal of the intein sequence by cis-splicing results in a functional GAT polypeptide. In another example, an encrypted GAT polypeptide can be expressed as two or more separate fragments; trans-splicing of these segments results in recovery of a functional GAT polypeptide. Various aspects of cis and trans splicing, gene encryption, and introduction of intervening sequences are described in more detail in US patent application Nos. 09/517,933 and 09/710,686, both of which are incorporated by reference herein in their entirety.

In general, the invention includes any polypeptide encoded by a modified GAT polynucleotide derived by mutation, recursive sequence recombination, and/or diversification of the polynucleotide sequences described herein. In some aspects of the invention, a GAT polypeptide is modified a by single or multiple amino acid substitution,

a deletion, an insertion, or a combination of one or more of these types of modifications. Substitutions can be conservative, or non-conservative, can alter function or not, and can add new function. Insertions and deletions can be substantial, such as the case of a truncation of a substantial fragment of the sequence, or in the fusion of additional sequence, either internally or at N or C terminal. In some embodiments of the invention, a GAT polypeptide is part of a fusion protein comprising a functional addition such as, for example, a secretion signal, a chloroplast transit peptide, a purification tag, or any of numerous other functional groups that will be apparent to the skilled artisan, and which are described in more detail elsewhere in this specification.

Polypeptides of the invention may contain one or more modified amino acid. The presence of modified amino acids may be advantageous in, for example, (a) increasing polypeptide in vivo half-life, (b) reducing or increasing polypeptide antigenicity, (c) increasing polypeptide storage stability. Amino acid(s) are modified, for example, co-translationally or post-translationally during recombinant production (e.g., N-linked glycosylation at N-X-S/T motifs during expression in mammalian cells) or modified by synthetic means.

Non-limiting examples of a modified amino acid include a glycosylated amino acid, a sulfated amino acid, a prenlyated (e.g., farnesylated, geranylgeranylated) amino acid, an acetylated amino acid, an acylated amino acid, a PEG-ylated amino acid, a biotinylated amino acid, a carboxylated amino acid, a phosphorylated amino acid, and the like. References adequate to guide one of skill in the modification of amino acids are replete throughout the literature. Example protocols are found in Walker (1998) Protein Protocols on CD-ROM Human Press, Towata, NJ.

Recombinant methods for producing and isolating GAT polypeptides of the invention are described herein. In addition to recombinant production, the polypeptides may be produced by direct peptide synthesis using solid-phase techniques (e.g., Stewart et al. (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield J (1963) J. Am. Chem. Soc. 85:2149-2154). Peptide synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.) in accordance with the instructions provided by the manufacturer. For example, subsequences may be chemically synthesized separately and combined using chemical methods to provide full-length GAT polypeptdides. Peptides can also be ordered from a variety of sources.

In another aspect of the invention, a GAT polypeptide of the invention is used to produce antibodies which have, e.g., diagnostic uses, for example, related to the activity, distribution, and expression of GAT polypeptides, for example, in various tissues of a transgenic plant.

GAT homologue polypeptides for antibody induction do not require biological activity; however, the polypeptide or oligopeptide must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least 10 amino acids, preferably at least 15 or 20 amino acids. Short stretches of a GAT polypeptide may be fused with another protein, such as keyhole limpet hemocyanin, and antibody produced against the chimeric molecule.

Methods of producing polyclonal and monoclonal antibodies are known to those of skill in the art, and many antibodies are available. *See,* e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY; Stites et al. (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and Kohler and Milstein (1975) Nature 256: 495-497. Other suitable techniques for antibody preparation include selection of libraries of recombinant antibodies in phage or similar vectors. *See*, Huse et al. (1989) Science 246: 1275-1281; and Ward, et al. (1989) Nature 341: 544-546. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 μM, preferably at least about 0.01 μM or better, and most typically and preferably, 0.001 μM or better.

Additional details antibody production and engineering techniques can be found in Borrebaeck (ed) (1995) Antibody Engineering, 2nd Edition Freeman and Company, NY (Borrebaeck); McCafferty et al. (1996) Antibody Engineering, A Practical Approach IRL at Oxford Press, Oxford, England (McCafferty), and Paul (1995) Antibody Engineering Protocols Humana Press, Towata, NJ (Paul).

Sequence Variations

GAT polypeptides of the present invention include conservatively modified variations of the sequences disclosed herein as SEQ ID NOS: 6-10 and 263-514. Such conservatively modified variations comprise substitutions, additions or deletions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less

than about 5%, more typically less than about 4%, 2%, or 1%) in any of SEQ ID NOS: 6-10 and 263-514.

For example, a conservatively modified variation (e.g., deletion) of the 146 amino acid polypeptide identified herein as SEQ ID NO:6 will have a length of at least 140 amino acids, preferably at least 141 amino acids, more preferably at least 144 amino acids, and still more preferably at least 146 amino acids, corresponding to a deletion of less than about 5%, 4%, 2% or about 1%, or less of the polypeptide sequence.

Another example of a conservatively modified variation (e.g., a "conservatively substituted variation") of the polypeptide identified herein as SEQ ID NO:6 will contain "conservative substitutions", according to the six substitution groups set forth in Table 2 (*infra*), in up to about 7 residues (i.e., less than about 5%) of the 146 amino acid polypeptide.

The GAT polypeptide sequence homologues of the invention, including conservatively substituted sequences, can be present as part of larger polypeptide sequences such as occur in a GAT polypeptide, in a GAT fusion with a signal sequence, e.g., a chloraplast targeting sequence, or upon the addition of one or more domains for purification of the protein (e.g., poly his segments, FLAG tag segments, etc.). In the latter case, the additional functional domains have little or no effect on the activity of the GAT portion of the protein, or where the additional domains can be removed by post synthesis processing steps such as by treatment with a protease.

Defining Polypeptides by Immunoreactivity

Because the polypeptides of the invention provide a new class of enzymes with a defined activity, i.e., the acetylation of glyphosate, the polypeptides also provide new structural features which can be recognized, e.g., in immunological assays. The generation of antisera which specifically binds the polypeptides of the invention, as well as the polypeptides which are bound by such antisera, are a feature of the invention.

The invention includes GAT polypeptides that specifically bind to or that are specifically immunoreactive with an antibody or antisera generated against an immunogen comprising an amino acid sequence selected from one or more of SEQ ID NO:6 to SEQ ID NO:10. To eliminate cross-reactivity with other GAT homologues, the antibody or antisera is subtracted with available related proteins, such as those represented by the proteins or peptides corresponding to GenBank accession numbers available as of the filing date of this application, and exemplified by CAA70664, Z99109 and Y09476.

Where the accession number corresponds to a nucleic acid, a polypeptide encoded by the nucleic acid is generated and used for antibody/antisera subtraction purposes. Figure 3 tabulates the relative identity between exemplary GAT polypeptides and the most closely related sequence available in Genbank, YitI. The function of native YitI has yet to be elucidated, but the enzyme has been shown to possess detectable GAT activity.

In one typical format, the immunoassay uses a polyclonal antiserum which was raised against one or more polypeptide comprising one or more of the sequences corresponding to one or more of SEQ ID NOS: 6-10 and 263-514, or a substantial subsequence thereof (i.e., at least about 30% of the full length sequence provided). The full set of potential polypeptide immunogens derived from SEQ ID NOS: 6-10 and 263-514 are collectively referred to below as "the immunogenic polypeptides." The resulting antisera is optionally selected to have low cross-reactivity against other related sequences and any such cross-reactivity is removed by immunoabsorbtion with one or more of the related sequences, prior to use of the polyclonal antiserum in the immunoassay.

In order to produce antisera for use in an immunoassay, one or more of the immunogenic polypeptides is produced and purified as described herein. For example, recombinant protein may be produced in a bacterial cell line. An inbred strain of mice (used in this assay because results are more reproducible due to the virtual genetic identity of the mice) is immunized with the immunogenic protein(s) in combination with a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (*see*, Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York, for a standard description of antibody generation, immunoassay formats and conditions that can be used to determine specific immunoreactivity). Alternatively, one or more synthetic or recombinant polypeptide derived from the sequences disclosed herein is conjugated to a carrier protein and used as an immunogen.

Polyclonal sera are collected and titered against the immunogenic polypeptide in an immunoassay, for example, a solid phase immunoassay with one or more of the immunogenic proteins immobilized on a solid support. Polyclonal antisera with a titer of 10^6 or greater are selected, pooled and subtracted with related polypeptides, e.g., those identified from GENBANK as noted, to produce subtracted pooled titered polyclonal antisera.

The subtracted pooled titered polyclonal antisera are tested for cross reactivity against the related polypeptides. Preferably at least two of the immunogenic GATs are used in this determination, preferably in conjunction with at least two of related

polypeptides, to identify antibodies which are specifically bound by the immunogenic protein(s).

In this comparative assay, discriminatory binding conditions are determined for the subtracted titered polyclonal antisera which result in at least about a 5-10 fold higher signal to noise ratio for binding of the titered polyclonal antisera to the immunogenic GAT polypeptides as compared to binding to the related polypeptides. That is, the stringency of the binding reaction is adjusted by the addition of non-specific competitors such as albumin or non-fat dry milk, or by adjusting salt conditions, temperature, or the like. These binding conditions are used in subsequent assays for determining whether a test polypeptide is specifically bound by the pooled subtracted polyclonal antisera. In particular, test polypeptides which show at least a 2-5x higher signal to noise ratio than the control polypeptides under discriminatory binding conditions, and at least about a ½ signal to noise ratio as compared to the immunogenic polypeptide(s), shares substantial structural similarity with the immunogenic polypeptide as compared to known GAT, and is, therefore a polypeptide of the invention.

In another example, immunoassays in the competitive binding format are used for detection of a test polypeptide. For example, as noted, cross-reacting antibodies are removed from the pooled antisera mixture by immunoabsorbtion with the control GAT polypeptides. The immunogenic polypeptide(s) are then immobilized to a solid support which is exposed to the subtracted pooled antisera. Test proteins are added to the assay to compete for binding to the pooled subtracted antisera. The ability of the test protein(s) to compete for binding to the pooled subtracted antisera as compared to the immobilized protein(s) is compared to the ability of the immunogenic polypeptide(s) added to the assay to compete for binding (the immunogenic polypeptides compete effectively with the immobilized immunogenic polypeptides for binding to the pooled antisera). The percent cross-reactivity for the test proteins is calculated, using standard calculations.

In a parallel assay, the ability of the control proteins to compete for binding to the pooled subtracted antisera is optionally determined as compared to the ability of the immunogenic polypeptide(s) to compete for binding to the antisera. Again, the percent cross-reactivity for the control polypeptides is calculated, using standard calculations. Where the percent cross-reactivity is at least 5-10x as high for the test polypeptides, the test polypeptides are said to specifically bind the pooled subtracted antisera.

In general, the immunoabsorbed and pooled antisera can be used in a competitive binding immunoassay as described herein to compare any test polypeptide to

the immunogenic polypeptide(s). In order to make this comparison, the two polypeptides are each assayed at a wide range of concentrations and the amount of each polypeptide required to inhibit 50% of the binding of the subtracted antisera to the immobilized protein is determined using standard techniques. If the amount of the test polypeptide required is less than twice the amount of the immunogenic polypeptide that is required, then the test polypeptide is said to specifically bind to an antibody generated to the immunogenic protein, provided the amount is at least about 5-10x as high as for a control polypeptide.

As a final determination of specificity, the pooled antisera is optionally fully immunosorbed with the *immunogenic* polypeptide(s) (rather than the control polypeptides) until little or no binding of the resulting immunogenic polypeptide subtracted pooled antisera to the immunogenic polypeptide(s) used in the immunosorbtion is detectable. This fully immunosorbed antisera is then tested for reactivity with the test polypeptide. If little or no reactivity is observed (i.e., no more than 2x the signal to noise ratio observed for binding of the fully immunosorbed antisera to the immunogenic polypeptide), then the test polypeptide is specifically bound by the antisera elicited by the immunogenic protein.

GLYPHOSATE N-ACETYLTRANSFERASE POLYNUCLEOTIDES

In one aspect, the invention provides a novel family of isolated or recombinant polynucleotides referred to herein as "glyphosate N-acetyltransferase polynucleotides" or "GAT polynucleotides." GAT polynucleotide sequences are characterized by the ability to encode a GAT polypeptide. In general, the invention includes any nucleotide sequence that encodes any of the novel GAT polypeptides described herein. In some aspects of the invention, a GAT polynucleotide that encodes a GAT polypeptide with GAT activity is preferred.

In one aspect, the GAT polynucleotides comprise recombinant or isolated forms of naturally occurring nucleic acids isolated from an organism, e,g, a bacterial strain. Exemplary GAT polynucleotides, e.g., SEQ ID NO:1 to SEQ ID NO:5, were discovered by expression cloning of sequences from *Bacillus* strains exhibiting GAT activity. Briefly, a collection of approximately 500 *Bacillus* and *Pseudomonas* strains were screened for native ability to N-acetylate glyphosate. Strains were grown in LB overnight, harvested by centrifugation, permeabilizied in dilute toluene, and then washed and resuspended in a reaction mix containing buffer, 5 mM glyphosate, and 200 μM acetyl-CoA. The cells were incubated in the reaction mix for between 1 and 48 hours, at which time an equal volume of methanol was added to the reaction. The cells were then

pelleted by centrifugation and the supernatant was filtered before analysis by parent ion mode mass spectrometry. The product of the reaction was positively identified as N-acetylglyphosate by comparing the mass spectrometry profile of the reaction mix to an N-acetylglyphosate standard as shown in Figure 2. Product detection was dependent on inclusion of both substrates (acetylCoA and glyphosate) and was abolished by heat denaturing the bacterial cells.

Individual GAT polynucleotides were then cloned from the identified strains by functional screening. Genomic DNA was prepared and partially digested with Sau3A1 enzyme. Fragments of approximately 4 Kb were cloned into an E. coli expression vector and transformed into electrocompetent E. coli. Individual clones exhibiting GAT activity were identified by mass spectrometry following a reaction as described previously except that the toluene wash was replaced by permeabilization with PMBS. Genomic fragments were sequenced and the putative GAT polypeptide-encoding open reading frame identified. Identity of the GAT gene was confirmed by expression of the open reading frame in E. coli and detection of high levels of N-acetylglyphosate produced from reaction mixtures.

In another aspect of the invention, GAT polynucleotides are produced by diversifying, e.g., recombining and/or mutating one or more naturally occurring, isolated, or recombinant GAT polynucleotides. As described in more detail elsewhere herein, it is often possible to generate diversified GAT polynucleotides encoding GAT polypeptides with superior functional attributes, e.g., increased catalytic function, increased stability, higher expression level, than a GAT polynucleotide used as a substrate or parent in the diversification process.

The polynucleotides of the invention have a variety of uses in, for example: recombinant production (i.e., expression) of the GAT polypeptides of the invention; as transgenes (e.g., to confer herbicide resistance in transgenic plants); as selectable markers for transformation and plasmid maintenance; as immunogens; as diagnostic probes for the presence of complementary or partially complementary nucleic acids (including for detection of natural GAT coding nucleic acids; as substrates for further diversity generation, e.g., recombination reactions or mutation reactions to produce new and/or improved GAT homologues, and the like.

It is important to note that certain specific, substantial and credible utilities of GAT polynucleotides do not require that the polynucleotide encode a polypeptide with substantial GAT activity. For example, GAT polynucleotides that do not encode active

enzymes can be valuable sources of parental polynucleotides for use in diversification procedures to arrive at GAT polynucleotide variants, or non-GAT polynucleotides, with desirable functional properties (e.g., high kcat or kcat/Km, low Km, high stability towards heat or other environmental factor, high transcription or translation rates, resistance to proteolytic cleavage, reducing antigenicity, etc.). For example, nucleotide sequences encoding protease variants with little or no detectable activity have been used as parent polynucleotides in DNA shuffling experiments to produce progeny encoding highly active proteases (Ness et al. (1999) Nature Biotechnology 17:893-96).

Polynucleotide sequences produced by diversity generation methods or recursive sequence recombination ("RSR") methods (*e.g.*, DNA shuffling) are a feature of the invention. Mutation and recombination methods using the nucleic acids described herein are a feature of the invention. For example, one method of the invention includes recursively recombining one or more nucleotide sequences of the invention as described above and below with one or more additional nucleotides. The recombining steps are optionally performed *in vivo, ex vivo, in silico* or *in vitro*. Said diversity generation or recursive sequence recombination produces at least one library of recombinant modified GAT polynucleotides. Polypeptides encoded by members of this library are included in the invention.

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least 20, 30, or 50 or more bases, which hybridize under stringent or highly stringent conditions to a GAT polynucleotide sequence. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted herein.

In accordance with the present invention, GAT polynucleotides, including nucleotide sequences that encode GAT poolypeptides, fragments of GAT polypeptides, related fusion proteins, or functional equivalents thereof, are used in recombinant DNA molecules that direct the expression of the GAT polypeptides in appropriate host cells, such as bacterial or plant cells. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can also be used to clone and express the GAT polynucleotides.

The invention provides GAT polynucleotides that encode transcription and/or translation product that are subsequently spliced to ultimately produce functional GAT polypeptides. Splicing can be accomplished in vitro or in vivo, and can involve cis

or trans splicing. The substrate for splicing can be polynucleotides (e.g., RNA transcripts) or polypeptides. An example of cis splicing of a polynucleotide is where an intron inserted into a coding sequence is removed and the two flanking exon regions are spliced to generate a GAT polypeptide encoding sequence. An example of trans splicing would be where a GAT polynucleotide is encrypted by separating the coding sequence into two or more fragments that can be separately transcribed and then spliced to form the full-length GAT encoding sequence. The use of a splicing enhancer sequence (which can be introduced into a construct of the invention) can facilitate splicing either in cis or trans. Cis and trans splicing of polypeptides are described in more detail elsehwhere herein. More detailed description of cis and trans splicing can be found in US patent application Nos. 09/517,933 and 09/710,686.

Thus, some GAT polynucleotides do not directly encode a full-length GAT polypeptide, but rather encode a fragment or fragments of a GAT polypeptide. These GAT polynucleotides can be used to express a functional GAT polypeptide through a mechanism involving splicing, where splicing can occur at the level of polynucleotide (e.g., intron/exon) and/or polypeptide (e.g., intein/extein). This can be useful, for example, in controlling expression of GAT activity, since functional GAT polypeptide will only be expressed if all required fragments are expressed in an environment that permits splicing processes to generate functional product. In another example, introduction of one or more insertion sequences into a GAT polynucleotide can facilitate recombination with a low homology polynucleotide; use of an intron or intein for the insertion sequence facilitates the removal of the intervening sequence, thereby restoring function of the encoded variant.

As will be understood by those of skill in the art, it can be advantageous to modify a coding sequence to enhance its expression in a particular host. The genetic code is redundant with 64 possible codons, but most organisms preferentially use a subset of these codons. The codons that are utilized most often in a species are called optimal codons, and those not utilized very often are classified as rare or low-usage codons (*see*, e.g., Zhang SP et al. (1991) *Gene* 105:61-72). Codons can be substituted to reflect the preferred codon usage of the host, a process sometimes called "codon optimization" or "controlling for species codon bias."

Optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host (*see also*, Murray, E. et al. (1989) *Nuc. Acids Res.* 17:477-508) can be prepared, for example, to increase the rate of translation or to produce

recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced from a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *S. cerevisiae* and mammals are UAA and UGA respectively. The preferred stop codon for monocotyledonous plants is UGA, whereas insects and E. coli prefer to use UAA as the stop codon (Dalphin ME et al. (1996) Nuc. Acids Res. 24: 216-218). Methodology for optimizing a nucleotide sequence for expression in a plant is provided, for example, in U.S. Patent No. 6,015,891, and references cited therein.

One embodiment of the invention includes a GAT polynucleotide having optimal codons for expression in a relevant host, e.g., a transgenic plant host. This is particularly desirable when a GAT polynucleotide of bacterial origin is introduced into a transgenic plant, e.g., to confer glyphosate resistance to the plant.

The polynucleotide sequences of the present invention can be engineered in order to alter a GAT polynucleotide for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the gene product. For example, alterations may be introduced using techniques that are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, alter glycosylation patterns, change codon preference, introduce splice sites, etc.

As described in more detail herein, the polynucleotides of the invention include sequences which encode novel GAT polypeptides and sequences complementary to the coding sequences, and novel fragments of coding sequence and complements thereof. The polynucleotides can be in the form of RNA or in the form of DNA, and include mRNA, cRNA, synthetic RNA and DNA, genomic DNA and cDNA. The polynucleotides can be double-stranded or single-stranded, and if single-stranded, can be the coding strand or the non-coding (anti-sense, complementary) strand. The polynucleotides optionally include the coding sequence of a GAT polypeptide (i) in isolation, (ii) in combination with additional coding sequence, so as to encode, e.g., a fusion protein, a pre-protein, a prepro-protein, or the like, (iii) in combination with non-coding sequences, such as introns or inteins, control elements such as a promoter, an enhancer, a terminator element, or 5' and/or 3' untranslated regions effective for expression of the coding sequence in a suitable host, and/or (iv) in a vector or host environment in which the GAT polynucleotide is a heterologous gene. Sequences can also be found in combination with typical compositional formulations of nucleic acids, including in the presence of carriers, buffers, adjuvants, excipients and the like.

Polynucleotides and oligonucleotides of the invention can be prepared by standard solid-phase methods, according to known synthetic methods. Typically, fragments of up to about 100 bases are individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form essentially any desired continuous sequence. For example, polynucleotides and oligonucleotides of the invention can be prepared by chemical synthesis using, e.g., the classical phosphoramidite method described by Beaucage et al. (1981) *Tetrahedron Letters* 22:1859-69, or the method described by Matthes et al. (1984) *EMBO J.* 3: 801-05., e.g., as is typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized, e.g., in an automatic DNA synthesizer, purified, annealed, ligated and cloned in appropriate vectors.

In addition, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (http://www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, CA) and many others. Similarly, peptides and antibodies can be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, Inc. (http://www.htibio.com), BMA Biomedicals Ltd (U.K.), Bio.Synthesis, Inc., and many others.

Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., Carruthers *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams *et al.*, *J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

General texts which describe molecular biological techniques useful herein, including mutagenesis, include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, volume 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook *et al.*, Molecular Cloning - A Laboratory Manual (2nd Ed.), volumes 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook"); and Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel")). Examples of techniques

sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qβ-replicase amplification and other RNA polymerase mediated techniques (*e.g.*, NASBA) are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, (1987) U.S. Patent No. 4,683,202; PCR Protocols A Guide to Methods and Applications (Innis *et al.*, eds.) Academic Press Inc. San Diego, CA (1990); Arnheim & Levinson (October 1, 1990) Chemical and Engineering News 36-47; The Journal Of NIH Research (1991) 3:81-94; Kwoh *et al.* (1989) Proc. Natl. Acad. Sci. USA 86:1173; Guatelli *et al.* (1990) Proc. Natl. Acad. Sci. USA 87:1874; Lomell *et al.* (1989) J. Clin. Chem. 35:1826; Landegren *et al.*, (1988) Science 241:1077-1080; Van Brunt (1990) Biotechnology 8:291-294; Wu and Wallace, (1989) Gene 4:560; Barringer *et al.* (1990) Gene 89:117, and Sooknanan and Malek (1995) Biotechnology 13:563-564. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in Cheng *et al.* (1994) Nature 369:684-685 and the references therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. *See*, Ausbel, Sambrook and Berger, *all supra*.

Sequence Variations

It will be appreciated by those skilled in the art that due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding GAT polypeptides of the invention may be produced, some of which bear substantial identity to the nucleic acid sequences explicitly disclosed herein.

Table 1
Codon Table

Amino acids			Codon					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

For instance, inspection of the codon table (Table 1) shows that codons AGA, AGG, CGA, CGC, CGG, and CGU all encode the amino acid arginine. Thus, at every position in the nucleic acids of the invention where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described above without altering the encoded polypeptide. It is understood that U in an RNA sequence corresponds to T in a DNA sequence.

Using, as an example, the nucleic acid sequence corresponding to nucleotides 1-15 of SEQ ID NO:1, ATG ATT GAA GTC AAA, a silent variation of this sequence includes AGT ATC GAG GTG AAG, both sequences which encode the amino acid sequence MIEVK, corresponding to amino acids 1-5 of SEQ ID NO:6.

Such "silent variations" are one species of "conservatively modified variations", discussed below. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified by standard techniques to encode a functionally identical polypeptide. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in any described sequence. The invention provides each and every possible variation of nucleic acid

sequence encoding a polypeptide of the invention that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code (e.g., as set forth in Table 1) as applied to the nucleic acid sequence encoding a GAT homologue polypeptide of the invention. All such variations of every nucleic acid herein are specifically provided and described by consideration of the sequence in combination with the genetic code. Any variant can be produced as noted herein.

A group of two or more different codons that, when translated in the same context, all encode the same amino acid, are referred to herein as "synonoumous codons." As described herein, in some aspects of the invention a GAT polynucleotide is engineered for optimized codon usage in a desired host organism, for example a plant host. The term "optimized" or "optimal" are not meant to be restricted to the very best possible combination of codons, but simple indicates that the coding sequence as a whole possesses an improved usage of codons relative to a precursor polynucleotide from which it was derived. Thus, in one aspect the invention provides a method for producing a GAT polynucleotide variant by replacing at least one parental codon in a nucleotide sequence with a synonomous codon that is preferentially used in a desired host organism, e.g., a plant, relative to the parental codon.

"Conservatively modified variations" or, simply, "conservative variations" of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or, where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. One of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 4%, 2% or 1%, or less) in an encoded sequence are "conservatively modified variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid.

Conservative substitution tables providing functionally similar amino acids are well known in the art. Table 2 sets forth six groups which contain amino acids that are "conservative substitutions" for one another.

Table 2
Conservative Substitution Groups

1	Alanine (A)	Serine (S)	Threonine (T)	
2	Aspartic acid (D)	Glutamic acid (E)		
3	Asparagine (N)	Glutamine (Q)		
4	Arginine (R)	Lysine (K)		
5	Isoleucine (I)	Leucine (L)	Methionine (M)	Valine (V)
6	Phenylalanine (F)	Tyrosine (Y)	Tryptophan (W)	

Thus, "conservatively substituted variations" of a listed polypeptide sequence of the present invention include substitutions of a small percentage, typically less than 5%, more typically less than 2% and often less than 1%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group.

For example, a conservatively substituted variation of the polypeptide identified herein as SEQ ID NO:6 will contain "conservative substitutions", according to the six groups defined above, in up to 7 residues (i.e., 5% of the amino acids) in the 146 amino acid polypeptide.

In a further example, if four conservative substitutions were localized in the region corresponding to amino acids 21 to 30 of SEQ ID NO:6, examples of conservatively substituted variations of this region,

RPN QPL EAC M, include:

KPQ QPV ESC M and

KPN NPL DAC V and the like, in accordance with the conservative substitutions listed in Table 2 (in the above example, conservative substitutions are underlined). Listing of a protein sequence herein, in conjunction with the above substitution table, provides an express listing of all conservatively substituted proteins.

Finally, the addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional or non-coding sequence, is a conservative variation of the basic nucleic acid.

One of skill will appreciate that many conservative variations of the nucleic acid constructs which are disclosed yield a functionally identical construct. For example, as discussed above, owing to the degeneracy of the genetic code, "silent substitutions"

(i.e., substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of *every* nucleic acid sequence which encodes an amino acid. Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the present invention.

Non-conservative modifications of a particular nucleic acid are those which substitute any amino acid not characterized as a conservative substitution. For example, any substitution which crosses the bounds of the six groups set forth in Table 2. These include substitutions of basic or acidic amino acids for neutral amino acids, (e.g., Asp, Glu, Asn, or Gln for Val, Ile, Leu or Met), aromatic amino acid for basic or acidic amino acids (e.g., Phe, Tyr or Trp for Asp, Asn, Glu or Gln) or any other substitution not replacing an amino acid with a like amino acid.

Nucleic Acid Hybridization

Nucleic acids "hybridize" when they associate, typically in solution. Nucleic acids hybridize due to a variety of well-characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, New York), as well as in Ausubel, *supra*, Hames and Higgins (1995) *Gene Probes 1,* IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 1) and Hames and Higgins (1995) *Gene Probes 2*, IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 2) provide details on the synthesis, labeling, detection and quantification of DNA and RNA, including oligonucleotides.

"Stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments, such as Southern and northern hybridizations, are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), *supra*, and in Hames and Higgins 1 and Hames and Higgins 2, *supra*.

For purposes of the present invention, generally, "highly stringent" hybridization and wash conditions are selected to be about 5°C or less lower than the

thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH (as noted below, highly stringent conditions can also be referred to in comparative terms). The T_m is the temperature (under defined ionic strength and pH) at which 50% of the test sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe.

The T_m of a nucleic acid duplex indicates the temperature at which the duplex is 50% denatured under the given conditions and its represents a direct measure of the stability of the nucleic acid hybrid. Thus, the T_m corresponds to the temperature corresponding to the midpoint in transition from helix to random coil; it depends on length, nucleotide composition, and ionic strength for long stretches of nucleotides.

After hybridization, unhybridized nucleic acid material can be removed by a series of washes, the stringency of which can be adjusted depending upon the desired results. Low stringency washing conditions (*e.g.*, using higher salt and lower temperature) increase sensitivity, but can product nonspecific hybridization signals and high background signals. Higher stringency conditions (*e.g.*, using lower salt and higher temperature that is closer to the hybridization temperature) lowers the background signal, typically with only the specific signal remaining. *See* Rapley, R. and Walker, J.M. eds., *Molecular Biomethods Handbook* (Humana Press, Inc. 1998) (hereinafter "Rapley and Walker"), which is incorporated herein by reference in its entirety for all purposes.

The T_m of a DNA-DNA duplex can be estimated using Equation 1 as follows:

$$T_m\ (°C) = 81.5°C + 16.6\ (\log_{10}M) + 0.41\ (\%G + C) - 0.72\ (\%f) - 500/n,$$

where M is the molarity of the monovalent cations (usually Na+), (%G + C) is the percentage of guanosine (G) and cystosine (C) nucleotides, (%f) is the percentage of formalize and n is the number of nucleotide bases (*i.e.*, length) of the hybrid. *See* Rapley and Walker, *supra*.

The T_m of an RNA-DNA duplex can be estimated by using Equation 2 as follows:

$$T_m\ (°C) = 79.8°C + 18.5\ (\log_{10}M) + 0.58\ (\%G + C) - 11.8(\%G + C)^2 - 0.56\ (\%f) - 820/n,$$

where M is the molarity of the monovalent cations (usually Na+), (%G + C)is the percentage of guanosine (G) and cystosine (C) nucleotides, (%f) is the percentage of formamide and n is the number of nucleotide bases (*i.e.*, length) of the hybrid. *Id.*

Equations 1 and 2 are typically accurate only for hybrid duplexes longer than about 100-200 nucleotides. *Id.*

The Tm of nucleic acid sequences shorter than 50 nucleotides can be calculated as follows:

T_m (°C) = 4(G + C) + 2(A + T),

where A (adenine), C, T (thymine), and G are the numbers of the corresponding nucleotides.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook, *supra* for a description of SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove background probe signal. An example low stringency wash is 2x SSC at 40°C for 15 minutes.

In general, a signal to noise ratio of 2.5x-5x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Detection of at least stringent hybridization between two sequences in the context of the present invention indicates relatively strong structural similarity or homology to, *e.g.*, the nucleic acids of the present invention provided in the sequence listings herein.

As noted, "highly stringent" conditions are selected to be about 5° C or less lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Target sequences that are closely related or identical to the nucleotide sequence of interest (*e.g.*, "probe") can be identified under highly stringent conditions. Lower stringency conditions are appropriate for sequences that are less complementary. *See, e.g.*, Rapley and Walker, *supra*.

Comparative hybridization can be used to identify nucleic acids of the invention, and this comparative hybridization method is a preferred method of distinguishing nucleic acids of the invention. Detection of highly stringent hybridization between two nucleotide sequences in the context of the present invention indicates relatively strong structural similarity/homology to, *e.g.*, the nucleic acids provided in the sequence listing herein. Highly stringent hybridization between two nucleotide sequences demonstrates a degree of similarity or homology of structure, nucleotide base composition, arrangement or order that is greater than that detected by stringent hybridization

conditions. In particular, detection of highly stringent hybridization in the context of the present invention indicates strong structural similarity or structural homology (*e.g.*, nucleotide structure, base composition, arrangement or order) to, *e.g.*, the nucleic acids provided in the sequence listings herein. For example, it is desirable to identify test nucleic acids that hybridize to the exemplar nucleic acids herein under stringent conditions.

Thus, one measure of stringent hybridization is the ability to hybridize to one of the listed nucleic acids (*e.g.*, nucleic acid sequences SEQ ID NO:1 to SEQ ID NO:5 and SEQ ID NO:11 to SEQ ID NO:262, and complementary polynucleotide sequences thereof), under highly stringent conditions (or very stringent conditions, or ultra-high stringency hybridization conditions, or ultra-ultra high stringency hybridization conditions). Stringent hybridization (as well as highly stringent, ultra-high stringency, or ultra-ultra high stringency hybridization conditions) and wash conditions can easily be determined empirically for any test nucleic acid. For example, in determining highly stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased (*e.g.*, by increasing temperature, decreasing salt concentration, increasing detergent concentration and/or increasing the concentration of organic solvents, such as formalin, in the hybridization or wash), until a selected set of criteria are met. For example, the hybridization and wash conditions are gradually increased until a probe comprising one or more nucleic acid sequences selected from SEQ ID NO:1 to SEQ ID NO:5 and SEQ ID NO:11 to SEQ ID NO:262, and complementary polynucleotide sequences thereof, binds to a perfectly matched complementary target (again, a nucleic acid comprising one or more nucleic acid sequences selected from SEQ ID NO:1 to SEQ ID NO:5 and SEQ ID NO:11 to SEQ ID NO:262, and complementary polynucleotide sequences thereof), with a signal to noise ratio that is at least about 2.5x, and optionally about 5x or more as high as that observed for hybridization of the probe to an unmatched target. In this case, the unmatched target is a nucleic acid corresponding to a nucleic acid (other than those in the accompanying sequence listing) that is present in a public database such as GenBank™ at the time of filing of the subject application. Such sequences can be identified in GenBank by one of skill. Examples include Accession Nos. Z99109 and Y09476. Additional such sequences can be identified in *e.g.*, GenBank, by one of ordinary skill in the art.

A test nucleic acid is said to specifically hybridize to a probe nucleic acid when it hybridizes at least ½ as well to the probe as to the perfectly matched

complementary target, i.e., with a signal to noise ratio at least ½ as high as hybridization of the probe to the target under conditions in which the perfectly matched probe binds to the perfectly matched complementary target with a signal to noise ratio that is at least about 2x-10x, and occasionally 20x, 50x or greater than that observed for hybridization to any of the unmatched polynucleotides Accession Nos. Z99109 and Y09476.

Ultra high-stringency hybridization and wash conditions are those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x as high as that observed for hybridization to any of the unmatched target nucleic acids Genbank Accession numbers Z99109 and Y09476. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-high stringency conditions.

Similarly, even higher levels of stringency can be determined by gradually increasing the hybridization and/or wash conditions of the relevant hybridization assay. For example, those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x, 20X, 50X, 100X, or 500X or more as high as that observed for hybridization to any of the unmatched target nucleic acids Genbank Accession numbers Z99109 and Y09476. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-ultra-high stringency conditions.

Target nucleic acids which hybridize to the nucleic acids represented by SEQ ID NO:1 to SEQ ID NO:5 and SEQ ID NO:11 to SEQ ID NO:262 under high, ultra-high and ultra-ultra high stringency conditions are a feature of the invention. Examples of such nucleic acids include those with one or a few silent or conservative nucleic acid substitutions as compared to a given nucleic acid sequence.

Nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code, or when antisera or antiserum generated against one or more of SEQ ID NO:6 to SEQ ID NO:10 and SEQ ID NO:263 to SEQ ID NO:514, which has been subtracted using the polypeptides encoded by known

nucleotide sequences, including Genbank Accession number CAA70664. Further details on immunological identification of polypeptides of the invention are found below. Additionally, for distinguishing between duplexes with sequences of less than about 100 nucleotides, a TMAC1 hybridization procedure known to those of ordinary skill in the art can be used. *See, e.g.*, Sorg, U. *et al.* 1 *Nucleic Acids Res.* (Sept. 11, 1991) 19(17), incorporated herein by reference in its entirety for all purposes.

In one aspect, the invention provides a nucleic acid which comprises a unique subsequence in a nucleic acid selected from SEQ ID NO:1 to SEQ ID NO:5 and SEQ ID NO:11 to SEQ ID NO:262. The unique subsequence is unique as compared to a nucleic acid corresponding to any of Genbank Accession numbers Z99109 and Y09476. Such unique subsequences can be determined by aligning any of SEQ ID NO:1 to SEQ ID NO:5 and SEQ ID NO:11 to SEQ ID NO:262 against the complete set of nucleic acids represented by GenBank accession numbers Z99109, Y09476 or other related sequences available in public databases as of the filing date of the subject application. Alignment can be performed using the BLAST algorithm set to default parameters. Any unique subsequence is useful, e.g., as a probe to identify the nucleic acids of the invention.

Similarly, the invention includes a polypeptide which comprises a unique subsequence in a polypeptide selected from: SEQ ID NO:6 to SEQ ID NO:10 and SEQ ID NO:263 to SEQ ID NO:514. Here, the unique subsequence is unique as compared to a polypeptide corresponding to GenBank accession number CAA70664. Here again, the polypeptide is aligned against the sequences represented by accession number CAA70664. Note that if the sequence corresponds to a non-translated sequence such as a pseudo gene, the corresponding polypeptide is generated simply by in silico translation of the nucleic acid sequence into an amino acid sequence, where the reading frame is selected to correspond to the reading frame of homologous GAT polynucleotides.

The invention also provides for target nucleic acids which hybridizes under stringent conditions to a unique coding oligonucleotide which encodes a unique subsequence in a polypeptide selected from SEQ ID NO:6 to SEQ ID NO:10 and SEQ ID NO:263 to SEQ ID NO:514, wherein the unique subsequence is unique as compared to a polypeptide corresponding to any of the control polypeptides. Unique sequences are determined as noted above.

In one example, the stringent conditions are selected such that a perfectly complementary oligonucleotide to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 2.5x-10x higher, preferably at least about a 5-10x

higher signal to noise ratio than for hybridization of the perfectly complementary oligonucleotide to a control nucleic acid corresponding to any of the control polypeptides. Conditions can be selected such that higher ratios of signal to noise are observed in the particular assay which is used, *e.g.*, about 15x, 20x, 30x, 50x or more. In this example, the target nucleic acid hybridizes to the unique coding oligonucleotide with at least a 2x higher signal to noise ratio as compared to hybridization of the control nucleic acid to the coding oligonucleotide. Again, higher signal to noise ratios can be selected, *e.g.*, about 2.5x, 5x, 10x, 20x, 30x, 50x or more. The particular signal will depend on the label used in the relevant assay, *e.g.*, a fluorescent label, a colorimetric label, a radioactive label, or the like.

Vectors, Promoters and Expression Systems.

The present invention also includes recombinant constructs comprising one or more of the nucleic acid sequences as broadly described above. The constructs comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts which describe molecular biological techniques useful herein, including the use of vectors, promoters and many other relevant topics, include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel"). Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qβ-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger, Sambrook, and Ausubel, as well as Mullis et al., (1987) U.S. Patent No. 4,683,202; PCR Protocols A Guide to Methods and Applications (Innis et al. eds)

Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3, 81-94; (Kwoh et al. (1989) Proc. Natl. Acad. Sci. USA 86, 1173; Guatelli et al. (1990) Proc. Natl. Acad. Sci. USA 87, 1874; Lomell et al. (1989) J. Clin. Chem 35, 1826; Landegren et al., (1988) Science 241, 1077-1080; Van Brunt (1990) Biotechnology 8, 291-294; Wu and Wallace, (1989) Gene 4, 560; Barringer et al. (1990) Gene 89, 117, and Sooknanan and Malek (1995) Biotechnology 13: 563-564. Improved methods for cloning in vitro amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. *See*, e.g., Ausubel, Sambrook and Berger, *all supra.*

The present invention also relates to engineered host cells that are transduced (transformed or transfected) with a vector of the invention (e.g., an invention cloning vector or an invention expression vector), as well as the production of polypeptides of the invention by recombinant techniques. The vector may be, for example, a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the GAT homologue gene. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, e.g., Sambrook, Ausubel and Berger, as well as e.g., Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley- Liss, New York and the references cited therein.

GAT polypeptides of the invention can be produced in non-animal cells such as plants, yeast, fungi, bacteria and the like. In addition to Sambrook, Berger and Ausubel, details regarding non-animal cell culture can be found in Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY; Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.

Polynucleotides of the present invention can be incorporated into any one of a variety of expression vectors suitable for expressing a polypeptide. Suitable vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated virus, retroviruses and many others. Any vector that transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host can be used.

When incorporated into an expression vector, a polynucleotide of the invention is operatively linked to an appropriate transcription control sequence (promoter) to direct mRNA synthesis. Examples of such transcription control sequences particularly suited for use in transgenic plants include the cauliflower mosaic virus (CaMV), figwort mosaic virus (FMV) and strawberry vein banding virus (SVBV) promoters, described in U.S. Provisional Application No. 60/245,354. Other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses and which can be used in some embodiments of the invention include SV40 promoter, *E. coli* lac or trp promoter, phage lambda P_L promoter. An expression vector optionally contains a ribosome binding site for translation initiation, and a transcription terminator. The vector also optionally includes appropriate sequences for amplifying expression, e.g., an enhancer. In addition, the expression vectors of the present invention optionally contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells, such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

Vectors of the present invention can be employed to transform an appropriate host to permit the host to express an invention protein or polypeptide. Examples of appropriate expression hosts include: bacterial cells, such as *E. coli*, *B. subtilis*, *Streptomyces*, and *Salmonella typhimurium*; fungal cells, such as *Saccharomyces cerevisiae, Pichia pastoris,* and *Neurospora crassa*; insect cells such as *Drosophila* and *Spodoptera frugiperda*; mammalian cells such as CHO, COS, BHK, HEK 293 or Bowes melanoma; or plant cells or explants, etc. It is understood that not all cells or cell lines need to be capable of producing fully functional GAT polypeptides; for example, antigenic fragments of a GAT polypeptide may be produced. The invention is not limited by the host cells employed.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the GAT polypeptide. For example, when large quantities of GAT polypeptide or fragments thereof are needed for commercial production or for induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified can be desirable. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the GAT polypeptide coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster (1989) J Biol Chem 264:5503-5509); pET vectors (Novagen, Madison WI); and the like.

Similarly, in the yeast *Saccharomyces cerevisiae* a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used for production of the GAT polypeptides of the invention. For reviews, see Ausubel et al. (*supra*) and Grant et al. (1987; Methods in Enzymology 153:516-544).

In mammalian host cells, a variety of expression systems, including viral-based systems, may be utilized. In cases where an adenovirus is used as an expression vector, a coding sequence, e.g., of a GAT polypeptide, is optionally ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion of a GAT polypeptide coding region into a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing a GAT in infected host cells (Logan and Shenk (1984) Proc Natl Acad Sci USA 81:3655-3659). In addition, transcription enhancers, such as the rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Similarly, in plant cells, expression can be driven from a transgene integrated into a plant chromosome, or cytoplasmically from an episomal or viral nucleic acid. In the case of stably integrated transgenes, it is often desirable to provide sequences capable of driving constitutive or inducible expression of the GAT polynucleotides of the invention, for example, using viral, e.g., CaMV, or plant derived regulatory sequences. Numerous plant derived regulatory sequences have been described, including sequences which direct expression in a tissue specific manner, e.g., TobRB7, patatin B33, GRP gene promoters, the rbcS-3A promoter, and the like. Alternatively, high level expression can be achieved by transiently expressing exogenous sequences of a plant viral vector, e.g., TMV, BMV, etc. Typically, transgenic plants constitutively expressing a GAT polynucleotide of

the invention will be preferred, and the regulatory sequences selected to insure constitutive stable expression of the GAT polypeptide.

In some embodiments of the present invention, a GAT polynucleotide construct suitable for transformation of plant cells is prepared. For example, a desired GAT polynucleotide can be incorporated into a recombinant expression cassette to facilitate introduction of the gene into a plant and subsequent expression of the encoded polypeptide. An expression cassette will typically comprise a GAT polynucleotide, or functional fragment thereof, operably linked to a promoter sequence and other transcriptional and translational initiation regulatory sequences which will direct expression of the sequence in the intended tissues (*e.g.*, entire plant, leaves, seeds) of the transformed plant.

For example, a strongly or weakly constitutive plant promoter can be employed which will direct expression of the GAT polypeptide all tissues of a plant. Such promoters are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciéns*, and other transcription initiation regions from various plant genes known to those of skill. In situations in which overexpression of a GAT poynucleotide is detrimental to the plant or otherwise undesirable, one of skill, upon review of this disclosure, will recognize that weak constitutive promoters can be used for low-levels of expression. In those cases where high levels of expression is not harmful to the plant, a strong promoter, e.g., a t-RNA or other pol III promoter, or a strong pol II promoter, such as the cauliflower mosaic virus promoter, can be used.

Alternatively, a plant promoter may be under environmental control. Such promoters are referred to here as "inducible" promoters. Examples of environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light.

The promoters used in the present invention can be "tissue-specific" and, as such, under developmental control in that the polynucleotide is expressed only in certain tissues, such as leaves and seeds. In embodiments in which one or more nucleic acid sequences endogenous to the plant system are incorporated into the construct, the endogenous promoters (or variants thereof) from these genes can be employed for directing expression of the genes in the transfected plant. Tissue-specific promoters can also be used to direct expression of heterologous polynucleotides.

In general, the particular promoter used in the expression cassette in plants depends on the intended application. Any of a number of promoters which direct transcription in plant cells are suitable. The promoter can be either constitutive or inducible. In addition to the promoters noted above, promoters of bacterial origin which operate in plants include the octopine synthase promoter, the nopaline synthase promoter and other promoters derived from native Ti plasmids (*see,* Herrara-Estrella *et al.* (1983) Nature 303:209-213). Viral promoters include the 35S and 19S RNA promoters of cauliflower mosaic virus (Odell *et al.* (1985) Nature 313:810-812). Other plant promoters include the ribulose-1,3-bisphosphate carboxylase small subunit promoter and the phaseolin promoter. The promoter sequence from the E8 gene and other genes may also be used. The isolation and sequence of the E8 promoter is described in detail in Deikman and Fischer (1988) EMBO J. 7:3315-3327.

To identify candidate promoters, the 5' portions of a genomic clone is analyzed for sequences characteristic of promoter sequences. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) as described by Messing *et al.* (1983) Genetic Engineering in Plants, Kosage, *et al.* (eds.), pp. 221-227.

In preparing polyucleotide constructs, e.g., vectors, of the invention, sequences other than the promoter and the cojoined polynucleotide can also be employed. If normal polypeptide expression is desired, a polyadenylation region at the 3'-end of a GAT-encoding region can be included. The polyadenylation region can be derived, for example, from a variety of plant genes, or from T-DNA.

The construct can also include a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide tolerance, particularly antibiotic tolerance, such as tolerance to kanamycin, G418, bleomycin, hygromycin, or herbicide tolerance, such as tolerance to chlorosluforon, or phosphinothricin (the active ingredient in the herbicides bialaphos and Basta).

Specific initiation signals can aid in efficient translation of a GAT polynucleotide-encoding sequence of the present invention. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a GAT polypeptide-encoding sequence, its initiation codon and upstream sequences are inserted into an appropriate expression vector, no additional translational control signals may be

needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf D et al. (1994) Results Probl Cell Differ 20:125-62; Bittner et al. (1987) Methods in Enzymol 153:516-544).

Secretion/Localization Sequences

Polynucleotides of the invention can also be fused, for example, in-frame to nucleic acids encoding a secretion/localization sequence, to target polypeptide expression to a desired cellular compartment, membrane, or organelle of a mammalian cell, or to direct polypeptide secretion to the periplasmic space or into the cell culture media. Such sequences are known to those of skill, and include secretion leader peptides, organelle targeting sequences (e.g., nuclear localization sequences, ER retention signals, mitochondrial transit sequences, chloroplast transit sequences), membrane localization/anchor sequences (e.g., stop transfer sequences, GPI anchor sequences), and the like.

In a preferred embodiment, a polynucleotide of the invention is fused in frame with an N-terminal chloroplast transit sequence (or chloroplast transit peptide sequence) derived from a gene encoding a polypeptide that is normally targeted to the chloroplast. Such sequences are typically rich in serine and threonine; are deficient in aspartate, glutamate, and tyrosine; and generally have a central domain rich in positively charged amino acids.

Expression Hosts

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a eukaryotic cell, such as a mammalian cell, a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, electroporation, or other common techniques (Davis, L., Dibner, M., and Battey, I. (1986) Basic Methods in Molecular Biology).

A host cell strain is optionally chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired

fashion. Such modifications of the protein include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing that cleaves a "pre" or a "prepro" form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as *E. coli, Bacillus sp.*, yeast or mammalian cells such as CHO, HeLa, BHK, MDCK, 293, WI38, etc. have specific cellular machinery and characteristic mechanisms, e.g., for post-translational activities and may be chosen to ensure the desired modification and processing of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression systems can be used. For example, plant cells, explants or tissues, e.g. shoots, leaf discs, which stably express a polypeptide of the invention are transduced using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for a period determined to be appropriate for the cell type, e.g., 1 or more hours for bacterial cells, 1-4 days for plant cells, 2-4 weeks for some plant explants, in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. For example, transgenic plants expressing the polypeptides of the invention can be selected directly for resistance to the herbicide, glyphosate. Resistant embryos derived from stably transformed explants can be proliferated, e.g., using tissue culture techniques appropriate to the cell type.

Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing GAT polynucleotides of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

Additional Polypeptide Sequences

Polynucleotides of the present invention may also comprise a coding sequence fused in-frame to a marker sequence that, e.g., facilitates purification of the encoded polypeptide. Such purification facilitating domains include, but are not limited

to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, a sequence which binds glutathione (e.g., GST), a hemagglutinin (HA) tag (corresponding to an epitope derived from the influenza hemagglutinin protein; Wilson et al. (1984) Cell 37:767), maltose binding protein sequences, the FLAG epitope utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle, WA), and the like. The inclusion of a protease-cleavable polypeptide linker sequence between the purification domain and the GAT homologue sequence is useful to facilitate purification. One expression vector contemplated for use in the compositions and methods described herein provides for expression of a fusion protein comprising a polypeptide of the invention fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography, as described in Porath et al. (1992) Protein Expression and Purification 3:263-281) while the enterokinase cleavage site provides a means for separating the GAT homologue polypeptide from the fusion protein. pGEX vectors (Promega; Madison, WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (e.g., glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

Polypeptide Production and Recovery

Following transduction of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, or other methods, which are well known to those skilled in the art.

As noted, many references are available for the culture and production of many cells, including cells of bacterial, plant, animal (especially mammalian) and archebacterial origin. *See* e.g., Sambrook, Ausubel, and Berger (*all supra*), as well as Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley- Liss, New York and the references cited therein; Doyle and Griffiths (1997) Mammalian Cell Culture: Essential Techniques John Wiley and Sons, NY; Humason (1979) Animal Tissue Techniques, fourth edition W.H. Freeman and Company; and

Ricciardelli, et al., (1989) In vitro Cell Dev. Biol. 25:1016-1024. For plant cell culture and regeneration, Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY; Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York); Jones, ed. (1984) Plant Gene Transfer and Expression Protocols, Humana Press, Totowa, New Jersey and Plant Molecular Biolgy (1993) R.R.D.Croy, Ed. Bios Scientific Publishers, Oxford, U.K. ISBN 0 12 198370 6. Cell culture media in general are set forth in Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL. Additional information for cell culture is found in available commercial literature such as the Life Science Research Cell Culture Catalogue (1998) from Sigma- Aldrich, Inc (St Louis, MO) ("Sigma-LSRCCC") and, e.g., The Plant Culture Catalogue and supplement (1997) also from Sigma-Aldrich, Inc (St Louis, MO) ("Sigma-PCCS"). Further details regarding plant cell transformation and transgenic plant production are found below.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by any of a number of methods well known in the art, including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography (e.g., using any of the tagging systems noted herein), hydroxylapatite chromatography, and lectin chromatography. Protein refolding steps can be used, as desired, in completing the configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed in the final purification steps. In addition to the references noted supra, a variety of purification methods are well known in the art, including, e.g., those set forth in Sandana (1997) Bioseparation of Proteins, Academic Press, Inc.; and Bollag et al. (1996) Protein Methods, 2nd Edition Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ, Harris and Angal (1990) Protein Purification Applications: A Practical Approach IRL Press at Oxford, Oxford, England; Harris and Angal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, England; Scopes (1993) Protein Purification: Principles and Practice 3rd Edition Springer Verlag, NY; Janson and Ryden (1998) Protein Purification: Principles, High Resolution Methods and Applications, Second Edition Wiley-VCH, NY; and Walker (1998) Protein Protocols on CD-ROM Humana Press, NJ.

In some cases, it is desirable to produce the GAT polypeptide of the invention in a large scale suitable for industrial and/or commercial applications. In such cases bulk fermentation procedures are employed. Briefly, a GAT polynucleotide, e.g., a polynucleotide comprising any one of SEQ ID NOS: 1-5 and 11-262. or other nucleic acids encoding GAT polypeptides of the invention can be cloned into an expression vector. For example, U.S. Patent No. 5,955,310 to Widner et al. "METHODS FOR PRODUCING A POLYPEPTIDE IN A BACILLUS CELL," describes a vector with tandem promoters, and stabilizing sequences operably linked to a polypeptide encoding sequence. After inserting the polynucleotide of interest into a vector, the vector is tranformed into a bacterial, e.g., a *Bacillus subtilis* strain PL1801IIE (amyE, apr, npr, spoIIE::Tn917) host. The introduction of an expression vector into a *Bacillus* cell may, for instance, be effected by protoplast transformation (see, e.g., Chang and Cohen (1979) Molecular General Genetics 168:111), by using competent cells (see, e.g., Young and Spizizin (1961) Journal of Bacteriology 81:823, or Dubnau and Davidoff-Abelson (1971) Journal of Molecular Biology 56:209), by electroporation (see, e.g., Shigekawa and Dower (1988) Biotechniques 6:742), or by conjugation (see, e.g., Koehler and Thorne (1987) Journal of Bacteriology 169:5271), also Ausubel, Sambrook and Berger, *all supra.*

The transformed cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods that are known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g., in catalogues of the American Type Culture Collection). The secreted polypeptide can be recovered directly from the medium.

The resulting polypeptide may be isolated by methods known in the art. For example, the polypeptide may be isolated from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation. The isolated polypeptide may then be further purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion),

electrophoretic procedures (e.g., preparative isoelectric focusing), differential solubility (e.g., ammonium sulfate precipitation), or extraction (see, e.g., Bollag et al. (1996) Protein Methods, 2nd Edition Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ; Bollag et al. (1996) Protein Methods, 2nd Edition Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ).

Cell-free transcription/translation systems can also be employed to produce polypeptides using DNAs or RNAs of the present invention. Several such systems are commercially available. A general guide to in vitro transcription and translation protocols is found in Tymms (1995) In vitro Transcription and Translation Protocols: Methods in Molecular Biology Volume 37, Garland Publishing, NY.

SUBSTRATES AND FORMATS FOR SEQUENCE RECOMBINATION

The polynucleotides of the invention are optionally used as substrates for a variety of diversity generating procedures, e.g., mutation, recombination and recursive recombination reactions, in addition to their use in standard cloning methods as set forth in, e.g., Ausubel, Berger and Sambrook, i.e., to produce additional GAT polynucleotides and polypeptides with desired properties. A variety of diversity generating protocols are available and described in the art. The procedures can be used separately, and/or in combination to produce one or more variants of a polynucleotide or set of polynucleotides, as well variants of encoded proteins. Individually and collectively, these procedures provide robust, widely applicable ways of generating diversified polynucleotides and sets of polynucleotides (including, e.g., polynucleotide libraries) useful, e.g., for the engineering or rapid evolution of polynucleotides, proteins, pathways, cells and/or organisms with new and/or improved characteristics. The process of altering the sequence can result in, for example, single nucleotide substitutions, multiple nucleotide substitutions, and insertion or deletion of regions of the nucleic acid sequence.

While distinctions and classifications are made in the course of the ensuing discussion for clarity, it will be appreciated that the techniques are often not mutually exclusive. Indeed, the various methods can be used singly or in combination, in parallel or in series, to access diverse sequence variants.

The result of any of the diversity generating procedures described herein can be the generation of one or more polynucleotides, which can be selected or screened for polynucleotides that encode proteins with or which confer desirable properties. Following diversification by one or more of the methods herein, or otherwise available to

one of skill, any polynucleotides that are produced can be selected for a desired activity or property, e.g. altered Km for glyphosate, altered Km for acetyl CoA, use of alternative cofactors (e.g., propionyl CoA) increased kcat, etc. This can include identifying any activity that can be detected, for example, in an automated or automatable format, by any of the assays in the art. For example, GAT homologs with increased specific activity can be detected by assaying the conversion of glyphosate to N-acetylglyphosate, e.g., by mass spectrometry. Alternatively, improved ability to confer resistance to glyphosate can be assayed by growing bacteria transformed with a nucleic acid of the invention on agar containing increasing concentrations of glyphosate or by spraying transgenic plants incorporating a nucleic acid of the invention with glyphosate. A variety of related (or even unrelated) properties can be evaluated, in serial or in parallel, at the discretion of the practitioner. Additional details regarding recombination and selection for herbicide tolerance can be found, e.g., in "DNA SHUFFLING TO PRODUCE HERBICIDE RESISTANT CROPS" (USSN 09/373,333) filed August 12,1999.

Descriptions of a variety of diversity generating procedures, including family shuffling and methods for generating modified nucleic acid sequences encoding multiple enzymatic domains, are found the following publications and the references cited therein: Soong, N. et al. (2000) "Molecular breeding of viruses" Nat Genet 25(4):436-39; Stemmer, et al. (1999) "Molecular breeding of viruses for targeting and other clinical properties" Tumor Targeting 4:1-4; Ness et al. (1999) "DNA Shuffling of subgenomic sequences of subtilisin" Nature Biotechnology 17:893-896; Chang et al. (1999) "Evolution of a cytokine using DNA family shuffling" Nature Biotechnology 17:793-797; Minshull and Stemmer (1999) "Protein evolution by molecular breeding" Current Opinion in Chemical Biology 3:284-290; Christians et al. (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" Nature Biotechnology 17:259-264; Crameri et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature 391:288-291; Crameri et al. (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," Nature Biotechnology 15:436-438; Zhang et al. (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" Proc. Natl. Acad. Sci. USA 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" Current Opinion in Biotechnology 8:724-733; Crameri et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" Nature Medicine 2:100-103; Crameri et al. (1996) "Improved green fluorescent protein by

molecular evolution using DNA shuffling" Nature Biotechnology 14:315-319; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'" Journal of Molecular Biology 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: The Encyclopedia of Molecular Biology. VCH Publishers, New York. pp.447-457; Crameri and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" BioTechniques 18:194-195; Stemmer et al., (1995) "Single-step assembly of a gene and entire plasmid form large numbers of oligodeoxy-ribonucleotides" Gene, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" Science 270: 1510; Stemmer (1995) "Searching Sequence Space" Bio/Technology 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by DNA shuffling" Nature 370:389-391; and Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution." Proc. Natl. Acad. Sci. USA 91:10747-10751.

Mutational methods of generating diversity include, for example, site-directed mutagenesis (Ling et al. (1997) "Approaches to DNA mutagenesis: an overview" Anal Biochem. 254(2): 157-178; Dale et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol. Biol. 57:369-374; Smith (1985) "In vitro mutagenesis" Ann. Rev. Genet. 19:423-462; Botstein & Shortle (1985) "Strategies and applications of in vitro mutagenesis" Science 229:1193-1201; Carter (1986) "Site-directed mutagenesis" Biochem. J. 237:1-7; and Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D.M.J. eds., Springer Verlag, Berlin)); mutagenesis using uracil containing templates (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Proc. Natl. Acad. Sci. USA 82:488-492; Kunkel et al. (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Methods in Enzymol. 154, 367-382; and Bass et al. (1988) "Mutant Trp repressors with new DNA-binding specificities" Science 242:240-245); oligonucleotide-directed mutagenesis (Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Zoller & Smith (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" Nucleic Acids Res. 10:6487-6500; Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" Methods in Enzymol. 100:468-500; and Zoller & Smith (1987) "Oligonucleotide-

directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" Methods in Enzymol. 154:329-350); phosphorothioate-modified DNA mutagenesis (Taylor et al. (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" Nucl. Acids Res. 13: 8749-8764; Taylor et al. (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" Nucl. Acids Res. 13: 8765-8787 (1985); Nakamaye & Eckstein (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" Nucl. Acids Res. 14: 9679-9698; Sayers et al. (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" Nucl. Acids Res. 16:791-802; and Sayers et al. (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" Nucl. Acids Res. 16: 803-814); mutagenesis using gapped duplex DNA (Kramer et al. (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" Nucl. Acids Res. 12: 9441-9456; Kramer & Fritz (1987) Methods in Enzymol. "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer et al. (1988) "Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" Nucl. Acids Res. 16: 7207; and Fritz et al. (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro" Nucl. Acids Res. 16: 6987-6999).

Additional suitable methods include point mismatch repair (Kramer et al. (1984) "Point Mismatch Repair" Cell 38:879-887), mutagenesis using repair-deficient host strains (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" Nucl. Acids Res. 13: 4431-4443; and Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" Methods in Enzymol. 154: 382-403), deletion mutagenesis (Eghtedarzadeh & Henikoff (1986) "Use of oligonucleotides to generate large deletions" Nucl. Acids Res. 14: 5115), restriction-selection and restriction-selection and restriction-purification (Wells et al. (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" Phil. Trans. R. Soc. Lond. A 317: 415-423), mutagenesis by total gene synthesis (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" Science 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" Nucl. Acids Res. 14:

6361-6372; Wells et al. (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" Gene 34:315-323; and Grundström et al. (1985) "Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis" Nucl. Acids Res. 13: 3305-3316), double-strand break repair (Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" Current Opinion in Biotechnology 4:450-455. "Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis" Proc. Natl. Acad. Sci. USA, 83:7177-7181). Additional details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

Additional details regarding various diversity generating methods can be found in the following U.S. patents, PCT publications, and EPO publications: U.S. Pat. No. 5,605,793 to Stemmer (February 25, 1997), "Methods for In Vitro Recombination;" U.S. Pat. No. 5,811,238 to Stemmer et al. (September 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" U.S. Pat. No. 5,830,721 to Stemmer et al. (November 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" U.S. Pat. No. 5,834,252 to Stemmer, et al. (November 10, 1998) "End-Complementary Polymerase Reaction;" U.S. Pat. No. 5,837,458 to Minshull, et al. (November 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" WO 95/22625, Stemmer and Crameri, "Mutagenesis by Random Fragmentation and Reassembly;" WO 96/33207 by Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction;" WO 97/20078 by Stemmer and Crameri "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" WO 97/35966 by Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" WO 99/41402 by Punnonen et al. "Targeting of Genetic Vaccine Vectors;" WO 99/41383 by Punnonen et al. "Antigen Library Immunization;" WO 99/41369 by Punnonen et al. "Genetic Vaccine Vector Engineering;" WO 99/41368 by Punnonen et al. "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008 by Stemmer and Crameri, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670 by Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" WO 99/23107 by Stemmer et al., "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" WO 99/21979 by Apt et al., "Human Papillomavirus Vectors;" WO 98/31837 by del Cardayre et al. "Evolution of Whole Cells and Organisms by

Recursive Sequence Recombination;" WO 98/27230 by Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" WO 98/13487 by Stemmer et al., "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection," WO 00/00632, "Methods for Generating Highly Diverse Libraries," WO 00/09679, "Methods for Obtaining in Vitro Recombined Polynucleotide Sequence Banks and Resulting Sequences," WO 98/42832 by Arnold et al., "Recombination of Polynucleotide Sequences Using Random or Defined Primers," WO 99/29902 by Arnold et al., "Method for Creating Polynucleotide and Polypeptide Sequences," WO 98/41653 by Vind, "An in Vitro Method for Construction of a DNA Library," WO 98/41622 by Borchert et al., "Method for Constructing a Library Using DNA Shuffling," and WO 98/42727 by Pati and Zarling, "Sequence Alterations using Homologous Recombination," WO 00/18906 by Patten et al., "Shuffling of Codon-Altered Genes;" WO 00/04190 by del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Recombination;" WO 00/42561 by Crameri et al., "Oligonucleotide Mediated Nucleic Acid Recombination;" WO 00/42559 by Selifonov and Stemmer "Methods of Populating Data Structures for Use in Evolutionary Simulations;" WO 00/42560 by Selifonov et al., "Methods for Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics;" WO 01/23401 by Welch et al., "Use of Codon-Varied Oligonucleotide Synthesis for Synthetic Shuffling;" and PCT/US01/06775 "Single-Stranded Nucleic Acid Template-Mediated Recombination and Nucleic Acid Fragment Isolation" by Affholter.

Certain U.S. applications provide additional details regarding various diversity generating methods, including "SHUFFLING OF CODON ALTERED GENES" by Patten et al. filed September 28, 1999, (USSN 09/407,800); "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION", by del Cardayre et al. filed July 15, 1998 (USSN 09/166,188), and July 15, 1999 (USSN 09/354,922); "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., filed September 28, 1999 (USSN 09/408,392), and "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., filed January 18, 2000 (PCT/US00/01203); "USE OF CODON-BASED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., filed September 28, 1999 (USSN 09/408,393); "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., filed January 18, 2000, (PCT/US00/01202) and, e.g., "METHODS FOR MAKING

CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., filed July 18, 2000 (USSN 09/618,579); "METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS" by Selifonov and Stemmer (PCT/US00/01138), filed January 18, 2000; and "SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION" by Affholter (USSN 60/186,482, filed March 2, 2000).

In brief, several different general classes of sequence modification methods, such as mutation, recombination, etc. are applicable to the present invention and set forth, e.g., in the references above. That is, alterations to the component nucleic acid sequences to produced modified gene fusion constructs can be performed by any number of the protocols described, either before cojoining of the sequences, or after the cojoining step. The following exemplify some of the different types of preferred formats for diversity generation in the context of the present invention, including, e.g., certain recombination based diversity generation formats.

Nucleic acids can be recombined in vitro by any of a variety of techniques discussed in the references above, including e.g., DNAse digestion of nucleic acids to be recombined followed by ligation and/or PCR reassembly of the nucleic acids. For example, sexual PCR mutagenesis can be used in which random (or pseudo random, or even non-random) fragmentation of the DNA molecule is followed by recombination, based on sequence similarity, between DNA molecules with different but related DNA sequences, in vitro, followed by fixation of the crossover by extension in a polymerase chain reaction. This process and many process variants is described in several of the references above, e.g., in Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751.

Similarly, nucleic acids can be recursively recombined in vivo, e.g., by allowing recombination to occur between nucleic acids in cells. Many such in vivo recombination formats are set forth in the references noted above. Such formats optionally provide direct recombination between nucleic acids of interest, or provide recombination between vectors, viruses, plasmids, etc., comprising the nucleic acids of interest, as well as other formats. Details regarding such procedures are found in the references noted above.

Whole genome recombination methods can also be used in which whole genomes of cells or other organisms are recombined, optionally including spiking of the genomic recombination mixtures with desired library components (e.g., genes

corresponding to the pathways of the present invention). These methods have many applications, including those in which the identity of a target gene is not known. Details on such methods are found, e.g., in WO 98/31837 by del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" and in, e.g., PCT/US99/15972 by del Cardayre et al., also entitled "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination." Thus, any of these processes and techniques for recombination, recursive recombination, and whole genome recombination, alone or in combination, can be used to generate the modified nucleic acid sequences and/or modified gene fusion constructs of the present invention.

Synthetic recombination methods can also be used, in which oligonucleotides corresponding to targets of interest are synthesized and reassembled in PCR or ligation reactions which include oligonucleotides which correspond to more than one parental nucleic acid, thereby generating new recombined nucleic acids. Oligonucleotides can be made by standard nucleotide addition methods, or can be made, e.g., by tri-nucleotide synthetic approaches. Details regarding such approaches are found in the references noted above, including, e.g., WO 00/42561 by Crameri et al., "Olgonucleotide Mediated Nucleic Acid Recombination;" WO 01/23401 by Welch et al., "Use of Codon-Varied Oligonucleotide Synthesis for Synthetic Shuffling;" WO 00/42560 by Selifonov et al., "Methods for Making Character Strings, Polynucleotides and Polypeptides Having Desired Characteristics;" and WO 00/42559 by Selifonov and Stemmer "Methods of Populating Data Structures for Use in Evolutionary Simulations."

In silico methods of recombination can be effected in which genetic algorithms are used in a computer to recombine sequence strings which correspond to homologous (or even non-homologous) nucleic acids. The resulting recombined sequence strings are optionally converted into nucleic acids by synthesis of nucleic acids which correspond to the recombined sequences, e.g., in concert with oligonucleotide synthesis/ gene reassembly techniques. This approach can generate random, partially random or designed variants. Many details regarding in silico recombination, including the use of genetic algorithms, genetic operators and the like in computer systems, combined with generation of corresponding nucleic acids (and/or proteins), as well as combinations of designed nucleic acids and/or proteins (e.g., based on cross-over site selection) as well as designed, pseudo-random or random recombination methods are described in WO 00/42560 by Selifonov et al., "Methods for Making Character Strings, Polynucleotides and Polypeptides Having Desired Characteristics" and WO 00/42559 by Selifonov and

Stemmer "Methods of Populating Data Structures for Use in Evolutionary Simulations." Extensive details regarding in silico recombination methods are found in these applications. This methodology is generally applicable to the present invention in providing for recombination of nucleic acid sequences and/or gene fusion constructs encoding proteins involved in various metabolic pathways (such as, for example, carotenoid biosynthetic pathways, ectoine biosynthetic pathways, polyhydroxyalkanoate biosynthetic pathways, aromatic polyketide biosynthetic pathways, and the like) in silico and/ or the generation of corresponding nucleic acids or proteins.

Many methods of accessing natural diversity, e.g., by hybridization of diverse nucleic acids or nucleic acid fragments to single-stranded templates, followed by polymerization and/or ligation to regenerate full-length sequences, optionally followed by degradation of the templates and recovery of the resulting modified nucleic acids can be similarly used. In one method employing a single-stranded template, the fragment population derived from the genomic library(ies) is annealed with partial, or, often approximately full length ssDNA or RNA corresponding to the opposite strand. Assembly of complex chimeric genes from this population is then mediated by nuclease-base removal of non-hybridizing fragment ends, polymerization to fill gaps between such fragments and subsequent single stranded ligation. The parental polynucleotide strand can be removed by digestion (e.g., if RNA or uracil-containing), magnetic separation under denaturing conditions (if labeled in a manner conducive to such separation) and other available separation/purification methods. Alternatively, the parental strand is optionally co-purified with the chimeric strands and removed during subsequent screening and processing steps. Additional details regarding this approach are found, e.g., in "Single-Stranded Nucleic Acid Template-Mediated Recombination and Nucleic Acid Fragment Isolation" by Affholter, PCT/US01/06775.

In another approach, single-stranded molecules are converted to double-stranded DNA (dsDNA) and the dsDNA molecules are bound to a solid support by ligand-mediated binding. After separation of unbound DNA, the selected DNA molecules are released from the support and introduced into a suitable host cell to generate a library enriched sequences which hybridize to the probe. A library produced in this manner provides a desirable substrate for further diversification using any of the procedures described herein.

Any of the preceding general recombination formats can be practiced in a reiterative fashion (e.g., one or more cycles of mutation/recombination or other diversity

generation methods, optionally followed by one or more selection methods) to generate a more diverse set of recombinant nucleic acids.

Mutagenesis employing polynucleotide chain termination methods have also been proposed (*see* e.g., U.S. Patent No. 5,965,408, "Method of DNA reassembly by interrupting synthesis" to Short, and the references above), and can be applied to the present invention. In this approach, double stranded DNAs corresponding to one or more genes sharing regions of sequence similarity are combined and denatured, in the presence or absence of primers specific for the gene. The single stranded polynucleotides are then annealed and incubated in the presence of a polymerase and a chain terminating reagent (e.g., ultraviolet, gamma or X-ray irradiation; ethidium bromide or other intercalators; DNA binding proteins, such as single strand binding proteins, transcription activating factors, or histones; polycyclic aromatic hydrocarbons; trivalent chromium or a trivalent chromium salt; or abbreviated polymerization mediated by rapid thermocycling; and the like), resulting in the production of partial duplex molecules. The partial duplex molecules, e.g., containing partially extended chains, are then denatured and reannealed in subsequent rounds of replication or partial replication resulting in polynucleotides which share varying degrees of sequence similarity and which are diversified with respect to the starting population of DNA molecules. Optionally, the products, or partial pools of the products, can be amplified at one or more stages in the process. Polynucleotides produced by a chain termination method, such as described above, are suitable substrates for any other described recombination format.

Diversity also can be generated in nucleic acids or populations of nucleic acids using a recombinational procedure termed "incremental truncation for the creation of hybrid enzymes" ("ITCHY") described in Ostermeier et al. (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" Nature Biotech 17:1205. This approach can be used to generate an initial a library of variants which can optionally serve as a substrate for one or more in vitro or in vivo recombination methods. See, also, Ostermeier et al. (1999) "Combinatorial Protein Engineering by Incremental Truncation," Proc. Natl. Acad. Sci. USA, 96: 3562-67; Ostermeier et al. (1999), "Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts," Biological and Medicinal Chemistry, 7: 2139-44.

Mutational methods which result in the alteration of individual nucleotides or groups of contiguous or non-contiguous nucleotides can be favorably employed to introduce nucleotide diversity into the nucleic acid sequences and/or gene fusion

constructs of the present invention. Many mutagenesis methods are found in the above-cited references; additional details regarding mutagenesis methods can be found in following, which can also be applied to the present invention.

For example, error-prone PCR can be used to generate nucleic acid variants. Using this technique, PCR is performed under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. Examples of such techniques are found in the references above and, e.g., in Leung et al. (1989) Technique 1:11-15 and Caldwell et al. (1992) PCR Methods Applic. 2:28-33. Similarly, assembly PCR can be used, in a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions can occur in parallel in the same reaction mixture, with the products of one reaction priming the products of another reaction.

Oligonucleotide directed mutagenesis can be used to introduce site-specific mutations in a nucleic acid sequence of interest. Examples of such techniques are found in the references above and, e.g., in Reidhaar-Olson et al. (1988) Science, 241:53-57. Similarly, cassette mutagenesis can be used in a process that replaces a small region of a double stranded DNA molecule with a synthetic oligonucleotide cassette that differs from the native sequence. The oligonucleotide can contain, e.g., completely and/or partially randomized native sequence(s).

Recursive ensemble mutagenesis is a process in which an algorithm for protein mutagenesis is used to produce diverse populations of phenotypically related mutants, members of which differ in amino acid sequence. This method uses a feedback mechanism to monitor successive rounds of combinatorial cassette mutagenesis. Examples of this approach are found in Arkin & Youvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815.

Exponential ensemble mutagenesis can be used for generating combinatorial libraries with a high percentage of unique and functional mutants. Small groups of residues in a sequence of interest are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Examples of such procedures are found in Delegrave & Youvan (1993) Biotechnology Research 11:1548-1552.

In vivo mutagenesis can be used to generate random mutations in any cloned DNA of interest by propagating the DNA, e.g., in a strain of E. coli that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a

higher random mutation rate than that of a wild-type parent. Propagating the DNA in one of these strains will eventually generate random mutations within the DNA. Such procedures are described in the references noted above.

Other procedures for introducing diversity into a genome, e.g. a bacterial, fungal, animal or plant genome can be used in conjunction with the above described and/or referenced methods. For example, in addition to the methods above, techniques have been proposed which produce nucleic acid multimers suitable for transformation into a variety of species (see, e.g., Schellenberger U.S. Patent No. 5,756,316 and the references above). Transformation of a suitable host with such multimers, consisting of genes that are divergent with respect to one another, (e.g., derived from natural diversity or through application of site directed mutagenesis, error prone PCR, passage through mutagenic bacterial strains, and the like), provides a source of nucleic acid diversity for DNA diversification, e.g., by an in vivo recombination process as indicated above.

Alternatively, a multiplicity of monomeric polynucleotides sharing regions of partial sequence similarity can be transformed into a host species and recombined in vivo by the host cell. Subsequent rounds of cell division can be used to generate libraries, members of which, include a single, homogenous population, or pool of monomeric polynucleotides. Alternatively, the monomeric nucleic acid can be recovered by standard techniques, e.g., PCR and/or cloning, and recombined in any of the recombination formats, including recursive recombination formats, described above.

Methods for generating multispecies expression libraries have been described (in addition to the reference noted above, see, e.g., Peterson et al. (1998) U.S. Pat. No. 5,783,431 "METHODS FOR GENERATING AND SCREENING NOVEL METABOLIC PATHWAYS," and Thompson, et al. (1998) U.S. Pat. No. 5,824,485 METHODS FOR GENERATING AND SCREENING NOVEL METABOLIC PATHWAYS) and their use to identify protein activities of interest has been proposed (In addition to the references noted above, see, Short (1999) U.S. Pat. No. 5,958,672 "PROTEIN ACTIVITY SCREENING OF CLONES HAVING DNA FROM UNCULTIVATED MICROORGANISMS"). Multispecies expression libraries include, in general, libraries comprising cDNA or genomic sequences from a plurality of species or strains, operably linked to appropriate regulatory sequences, in an expression cassette. The cDNA and/or genomic sequences are optionally randomly ligated to further enhance diversity. The vector can be a shuttle vector suitable for transformation and expression in more than one species of host organism, e.g., bacterial species, eukaryotic cells. In some

cases, the library is biased by preselecting sequences which encode a protein of interest, or which hybridize to a nucleic acid of interest. Any such libraries can be provided as substrates for any of the methods herein described.

The above described procedures have been largely directed to increasing nucleic acid and/ or encoded protein diversity. However, in many cases, not all of the diversity is useful, e.g., functional, and contributes merely to increasing the background of variants that must be screened or selected to identify the few favorable variants. In some applications, it is desirable to preselect or prescreen libraries (e.g., an amplified library, a genomic library, a cDNA library, a normalized library, etc.) or other substrate nucleic acids prior to diversification, e.g., by recombination-based mutagenesis procedures, or to otherwise bias the substrates towards nucleic acids that encode functional products. For example, in the case of antibody engineering, it is possible to bias the diversity generating process toward antibodies with functional antigen binding sites by taking advantage of in vivo recombination events prior to manipulation by any of the described methods. For example, recombined CDRs derived from B cell cDNA libraries can be amplified and assembled into framework regions (e.g., Jirholt et al. (1998) "Exploiting sequence space: shuffling in vivo formed complementarity determining regions into a master framework" Gene 215: 471) prior to diversifying according to any of the methods described herein.

Libraries can be biased towards nucleic acids which encode proteins with desirable enzyme activities. For example, after identifying a clone from a library which exhibits a specified activity, the clone can be mutagenized using any known method for introducing DNA alterations. A library comprising the mutagenized homologues is then screened for a desired activity, which can be the same as or different from the initially specified activity. An example of such a procedure is proposed in Short (1999) U.S. Patent No. 5,939,250 for "PRODUCTION OF ENZYMES HAVING DESIRED ACTIVITIES BY MUTAGENESIS." Desired activities can be identified by any method known in the art. For example, WO 99/10539 proposes that gene libraries can be screened by combining extracts from the gene library with components obtained from metabolically rich cells and identifying combinations which exhibit the desired activity. It has also been proposed (e.g., WO 98/58085) that clones with desired activities can be identified by inserting bioactive substrates into samples of the library, and detecting bioactive fluorescence corresponding to the product of a desired activity using a fluorescent analyzer, e.g., a flow cytometry device, a CCD, a fluorometer, or a spectrophotometer.

Libraries can also be biased towards nucleic acids which have specified characteristics, e.g., hybridization to a selected nucleic acid probe. For example, application WO 99/10539 proposes that polynucleotides encoding a desired activity (e.g., an enzymatic activity, for example: a lipase, an esterase, a protease, a glycosidase, a glycosyl transferase, a phosphatase, a kinase, an oxygenase, a peroxidase, a hydrolase, a hydratase, a nitrilase, a transaminase, an amidase or an acylase) can be identified from among genomic DNA sequences in the following manner. Single stranded DNA molecules from a population of genomic DNA are hybridized to a ligand-conjugated probe. The genomic DNA can be derived from either a cultivated or uncultivated microorganism, or from an environmental sample. Alternatively, the genomic DNA can be derived from a multicellular organism, or a tissue derived therefrom. Second strand synthesis can be conducted directly from the hybridization probe used in the capture, with or without prior release from the capture medium or by a wide variety of other strategies known in the art. Alternatively, the isolated single-stranded genomic DNA population can be fragmented without further cloning and used directly in, e.g., a recombination-based approach, that employs a single-stranded template, as described above.

"Non-Stochastic" methods of generating nucleic acids and polypeptides are alleged in Short "Non-Stochastic Generation of Genetic Vaccines and Enzymes" WO 00/46344. These methods, including proposed non-stochastic polynucleotide reassembly and site-saturation mutagenesis methods be applied to the present invention as well. Random or semi-random mutagenesis using doped or degenerate oligonucleotides is also described in, e.g., Arkin and Youvan (1992) "Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis" *Biotechnology* 10:297-300; Reidhaar-Olson et al. (1991) "Random mutagenesis of protein sequences using oligonucleotide cassettes" *Methods Enzymol.* 208:564-86; Lim and Sauer (1991) "The role of internal packing interactions in determining the structure and stability of a protein" *J. Mol. Biol.* 219:359-76; Breyer and Sauer (1989) "Mutational analysis of the fine specificity of binding of monoclonal antibody 51F to lambda repressor" *J. Biol. Chem.* 264:13355-60); and "Walk-Through Mutagenesis" (Crea, R; US Patents 5,830,650 and 5,798,208, and EP Patent 0527809 B1.

It will readily be appreciated that any of the above described techniques suitable for enriching a library prior to diversification can also be used to screen the products, or libraries of products, produced by the diversity generating methods. Any of

the above described methods can be practiced recursively or in combination to alter nucleic acids, e.g., GAT encoding polynucleotides.

Kits for mutagenesis, library construction and other diversity generation methods are also commercially available. For example, kits are available from, e.g., Stratagene (e.g., QuickChange™ site-directed mutagenesis kit; and Chameleon™ double-stranded, site-directed mutagenesis kit), Bio/Can Scientific, Bio-Rad (e.g., using the Kunkel method described above), Boehringer Mannheim Corp., Clonetech Laboratories, DNA Technologies, Epicentre Technologies (e.g., 5 prime 3 prime kit); Genpak Inc, Lemargo Inc, Life Technologies (Gibco BRL), New England Biolabs, Pharmacia Biotech, Promega Corp., Quantum Biotechnologies, Amersham International plc (e.g., using the Eckstein method above), and Anglian Biotechnology Ltd (e.g., using the Carter/Winter method above).

The above references provide many mutational formats, including recombination, recursive recombination, recursive mutation and combinations or recombination with other forms of mutagenesis, as well as many modifications of these formats. Regardless of the diversity generation format that is used, the nucleic acids of the present invention can be recombined (with each other, or with related (or even unrelated) sequences) to produce a diverse set of recombinant nucleic acids for use in the gene fusion constructs and modified gene fusion constructs of the present invention, including, e.g., sets of homologous nucleic acids, as well as corresponding polypeptides.

Many of the above-described methodologies for generating modified polynucleotides generate a large number of diverse variants of a parental sequence or sequences. In some preferred embodiments of the invention the modification technique (e.g., some form of shuffling) is used to generate a library of variants that is then screened for a modified polynucleotide or pool of modified polynucleotides encoding some desired functional attribute, e.g., improved GAT activity. Exemplary enzymatic activities that can be screened for include catalytic rates (conventionally characterized in terms of kinetic constants such as k_{cat} and K_M), substrate specificity, and susceptibility to activation or inhibition by substrate, product or other molecules (e.g., inhibitors or activators).

One example of selection for a desired enzymatic activity entails growing host cells under conditions that inhibit the growth and/or survival of cells that do not sufficiently express an enzymatic activity of interest, e.g. the GAT activity. Using such a selection process can eliminate from consideration all modified polynucleotides except those encoding a desired enzymatic activity. For example, in some embodiments of the

invention host cells are maintained under conditions that inhibit cell growth or survival in the absence of sufficient levels of GAT, e.g., a concentration of glyphosate that is lethal or inhibits the growth of a wild-type plant of the same variety that lack does not express GAT polynucleotide. Under these conditions, only a host cell harboring a modified nucleic acid that encodes enzymatic activity or activities able to catalyze production of sufficient levels of the product will survive and grow. Some embodiments of the invention employ multiples rounds of screening at increasing concentrations of glyphosate or a glyphosate analog.

In some embodiments of the invention, mass spectrometry is used to detect the acetylation of glyphosate, or a glyphosate analog or metabolite. The used of mass spectrometry is described in more detail in the Examples below.

For convenience and high throughput it will often be desirable to screen/select for desired modified nucleic acids in a microorganism, e.g., a bacteria such as *E. coli*. On the other hand, screening in plant cells or plants can will in some cases be preferable where the ultimate aim is to generate a modified nucleic acid for expression in a plant system.

In some preferred embodiments of the invention throughput is increased by screening pools of host cells expressing different modified nucleic acids, either alone or as part of a gene fusion construct. Any pools showing significant activity can be deconvoluted to identify single clones expressing the desirable activity.

The skilled artisan will recognize that the relevant assay, screening or selection method will vary depending upon the desired host organism, etc. It is normally advantageous to employ an assay that can be practiced in a high-throughput format.

In high through put assays, it is possible to screen up to several thousand different variants in a single day. For example, each well of a microtiter plate can be used to run a separate assay, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single variant.

In addition to fluidic approaches, it is possible, as mentioned above, simply to grow cells on media plates that select for the desired enzymatic or metabolic function. This approach offers a simple and high-throughput screening method.

A number of well known robotic systems have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations like the automated synthesis apparatus developed by Takeda Chemical Industries, LTD. (Osaka, Japan) and many robotic systems utilizing robotic arms (Zymate

II, Zymark Corporation, Hopkinton, MA.; Orca, Hewlett-Packard, Palo Alto, CA) which mimic the manual synthetic operations performed by a scientist. Any of the above devices are suitable for application to the present invention. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein with reference to the integrated system will be apparent to persons skilled in the relevant art.

High throughput screening systems are commercially available (*see*, e.g., Zymark Corp., Hopkinton, MA; Air Technical Industries, Mentor, OH; Beckman Instruments, Inc. Fullerton, CA; Precision Systems, Inc., Natick, MA, etc.). These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization.

The manufacturers of such systems provide detailed protocols for the various high throughput devices. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. Microfluidic approaches to reagent manipulation have also been developed, e.g., by Caliper Technologies (Mountain View, CA).

Optical images viewed (and, optionally, recorded) by a camera or other recording device (e.g., a photodiode and data storage device) are optionally further processed in any of the embodiments herein, e.g., by digitizing the image and/or storing and analyzing the image on a computer. A variety of commercially available peripheral equipment and software is available for digitizing, storing and analyzing a digitized video or digitized optical image, e.g., using PC (Intel x86 or pentium chip compatible DOS™, OS™ WINDOWS™, WINDOWS NT™ or WINDOWS 95™ based machines), MACINTOSH™, or UNIX based (e.g., SUN™ work station) computers.

One conventional system carries light from the assay device to a cooled charge-coupled device (CCD) camera, a common use in the art. A CCD camera includes an array of picture elements (pixels). The light from the specimen is imaged on the CCD. Particular pixels corresponding to regions of the specimen (e.g., individual hybridization sites on an array of biological polymers) are sampled to obtain light intensity readings for each position. Multiple pixels are processed in parallel to increase speed. The apparatus and methods of the invention are easily used for viewing any sample, e.g. by fluorescent or dark field microscopic techniques.

OTHER POLYNUCLEOTIDE COMPOSITIONS

The invention also includes compositions comprising two or more polynucleotides of the invention (e.g., as substrates for recombination). The composition can comprise a library of recombinant nucleic acids, where the library contains at least 2, 3, 5, 10, 20, or 50 or more polynucleotides. The polynucleotides are optionally cloned into expression vectors, providing expression libraries.

The invention also includes compositions produced by digesting one or more polynucleotide of the invention with a restriction endonuclease, an RNAse, or a DNAse (e.g., as is performed in certain of the recombination formats noted above); and compositions produced by fragmenting or shearing one or more polynucleotide of the invention by mechanical means (e.g., sonication, vortexing, and the like), which can also be used to provide substrates for recombination in the methods above. Similarly, compositions comprising sets of oligonucleotides corresponding to more than one nucleic acid of the invention are useful as recombination substrates and are a feature of the invention. For convenience, these fragmented, sheared, or oligonucleotide synthesized mixtures are referred to as fragmented nucleic acid sets.

Also included in the invention are compositions produced by incubating one or more of the fragmented nucleic acid sets in the presence of ribonucleotide- or deoxyribonucelotide triphosphates and a nucleic acid polymerase. This resulting composition forms a recombination mixture for many of the recombination formats noted above. The nucleic acid polymerase may be an RNA polymerase, a DNA polymerase, or an RNA-directed DNA polymerase (e.g., a "reverse transcriptase"); the polymerase can be, e.g., a thermostable DNA polymerase (such as, VENT, TAQ, or the like).

INTEGRATED SYSTEMS

The present invention provides computers, computer readable media and integrated systems comprising character strings corresponding to the sequence information herein for the polypeptides and nucleic acids herein, including, e.g., those sequences listed herein and the various silent substitutions and conservative substitutions thereof.

For example, various methods and genetic algorithms (GAs) known in the art can be used to detect homology or similarity between different character strings, or can be used to perform other desirable functions such as to control output files, provide the basis for making presentations of information including the sequences and the like. Examples include BLAST, discussed *supra*.

Thus, different types of homology and similarity of various stringency and length can be detected and recognized in the integrated systems herein. For example, many homology determination methods have been designed for comparative analysis of sequences of biopolymers, for spell-checking in word processing, and for data retrieval from various databases. With an understanding of double-helix pair-wise complement interactions among 4 principal nucleobases in natural polynucleotides, models that simulate annealing of complementary homologous polynucleotide strings can also be used as a foundation of sequence alignment or other operations typically performed on the character strings corresponding to the sequences herein (e.g., word-processing manipulations, construction of figures comprising sequence or subsequence character strings, output tables, etc.). An example of a software package with GAs for calculating sequence similarity is BLAST, which can be adapted to the present invention by inputting character strings corresponding to the sequences herein.

Similarly, standard desktop applications such as word processing software (e.g., Microsoft Word™ or Corel WordPerfect™) and database software (e.g., spreadsheet software such as Microsoft Excel™, Corel Quattro Pro™, or database programs such as Microsoft Access™ or Paradox™) can be adapted to the present invention by inputting a character string corresponding to the GAT homologues of the invention (either nucleic acids or proteins, or both). For example, the integrated systems can include the foregoing software having the appropriate character string information, e.g., used in conjunction with a user interface (e.g., a GUI in a standard operating system such as a Windows, Macintosh or LINUX system) to manipulate strings of characters. As noted, specialized alignment programs such as BLAST can also be incorporated into the systems of the invention for alignment of nucleic acids or proteins (or corresponding character strings).

Integrated systems for analysis in the present invention typically include a digital computer with GA software for aligning sequences, as well as data sets entered into the software system comprising any of the sequences herein. The computer can be, e.g., a PC (Intel x86 or Pentium chip- compatible DOS™, OS2™ WINDOWS™ WINDOWS NT™, WINDOWS95™, WINDOWS98™ LINUX based machine, a MACINTOSH™, Power PC, or a UNIX based (e.g., SUN™ work station) machine) or other commercially common computer which is known to one of skill. Software for aligning or otherwise manipulating sequences is available, or can easily be constructed by one of skill using a standard programming language such as Visualbasic, Fortran, Basic, Java, or the like.

Any controller or computer optionally includes a monitor which is often a cathode ray tube ("CRT") display, a flat panel display (e.g., active matrix liquid crystal display, liquid crystal display), or others. Computer circuitry is often placed in a box which includes numerous integrated circuit chips, such as a microprocessor, memory, interface circuits, and others. The box also optionally includes a hard disk drive, a floppy disk drive, a high capacity removable drive such as a writeable CD-ROM, and other common peripheral elements. Inputting devices such as a keyboard or mouse optionally provide for input from a user and for user selection of sequences to be compared or otherwise manipulated in the relevant computer system.

The computer typically includes appropriate software for receiving user instructions, either in the form of user input into a set parameter fields, e.g., in a GUI, or in the form of preprogrammed instructions, e.g., preprogrammed for a variety of different specific operations. The software then converts these instructions to appropriate language for instructing the operation of the fluid direction and transport controller to carry out the desired operation.

The software can also include output elements for controlling nucleic acid synthesis (e.g., based upon a sequence or an alignment of a sequences herein) or other operations which occur downstream from an alignment or other operation performed using a character string corresponding to a sequence herein. Nucleic acid synthesis equipment can, accordingly, be a component in one or more integrated systems herein.

In an additional aspect, the present invention provides kits embodying the methods, composition, systems and apparatus herein. Kits of the invention optionally comprise one or more of the following: (1) an apparatus, system, system component or apparatus component as described herein; (2) instructions for practicing the methods described herein, and/or for operating the apparatus or apparatus components herein and/or for using the compositions herein; (3) one or more GAT composition or component; (4) a container for holding components or compositions, and, (5) packaging materials.

In a further aspect, the present invention provides for the use of any apparatus, apparatus component, composition or kit herein, for the practice of any method or assay herein, and/or for the use of any apparatus or kit to practice any assay or method herein.

HOST CELLS AND ORGANISMS

The host cell can be eukaryotic, for example, a eukaryotic cell, a plant cell, an animal cell, a protoplast, or a tissue culture. The host cell optionally comprises a plurality of cells, for example, an organism. Alternatively, the host cell can be prokaryotic including, but not limited to, bacteria (i.e., gram positive bacteria, purple bacteria, green sulfur bacteria, green non-sulfur bacteria, cyanobacteria, spirochetes, thermatogales, flavobacteria, and bacteroides) and archaebacteria (i.e., Korarchaeota, Thermoproteus, Pyrodictium, Thermococcales, methanogens, Archaeoglobus, and extreme halophiles).

Transgenic plants, or plant cells, incorporating the GAT nucleic acids, and/or expressing the GAT polypeptides of the invention are a feature of the invention. The transformation of plant cells and protoplasts can be carried out in essentially any of the various ways known to those skilled in the art of plant molecular biology, including, but not limited to, the methods described herein. *See*, in general, Methods in Enzymology, Vol. 153 (*Recombinant DNA Part D*) Wu and Grossman (eds.) 1987, Academic Press, incorporated herein by reference. As used herein, the term "transformation" means alteration of the genotype of a host plant by the introduction of a nucleic acid sequence, e.g., a "heterologous" or "foreign" nucleic acid sequence. The heterologous nucleic acid sequence need not necessarily originate from a different source but it will, at some point, have been external to the cell into which is introduced.

In addition to Berger, Ausubel and Sambrook, useful general references for plant cell cloning, culture and regeneration include Jones (ed) (1995) Plant Gene Transfer and Expression Protocols-- Methods in Molecular Biology, Volume 49 Humana Press Towata NJ; Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY (Payne); and Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) (Gamborg). A variety of cell culture media are described in Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL (Atlas). Additional information for plant cell culture is found in available commercial literature such as the Life Science Research Cell Culture Catalogue (1998) from Sigma- Aldrich, Inc (St Louis, MO) (Sigma-LSRCCC) and, e.g., the Plant Culture Catalogue and supplement (1997) also from Sigma-Aldrich, Inc (St Louis, MO) (Sigma-PCCS). Additional details regarding plant cell culture are found in Croy, (ed.) (1993) Plant Molecular Biology Bios Scientific Publishers, Oxford, U.K.

In an embodiment of this invention, recombinant vectors including one or more GAT polynucleotides, suitable for the transformation of plant cells are prepared. A DNA sequence encoding for the desired GAT polypeptide, e.g., selected from among SEQ ID NOS: 1-5 and 11-262, is conveniently used to construct a recombinant expression cassette which can be introduced into the desired plant. In the context of the present invention, an expression cassette will typically comprise a selected GAT polynucleotide operably linked to a promoter sèquence and other transcriptional and translational initiation regulatory sequences which are sufficient to direct the transcription of the GAT sequence in the intended tissues (e.g., entire plant, leaves, roots, etc.) of the transformed plant.

For example, a strongly or weakly constitutive plant promoter that directs expression of a GAT nucleic acid in all tissues of a plant can be favorably employed. Such promoters are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the 1´- or 2´- promoter of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill. Where overexpression of a GAT polypeptide of the invention is detrimental to the plant, one of skill, will recognize that weak constitutive promoters can be used for low-levels of expression. In those cases where high levels of expression is not harmful to the plant, a strong promoter, e.g., a t-RNA, or other pol III promoter, or a strong pol II promoter, (e.g., the cauliflower mosaic virus promoter, CaMV, 35S promoter) can be used.

Alternatively, a plant promoter can be under environmental control. Such promoters are referred to as "inducible" promoters. Examples of environmental conditions that may alter transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light. In some cases, it is desirable to use promoters that are "tissue-specific" and/or are under developmental control such that the GAT polynucleotide is expressed only in certain tissues or stages of development, e.g., leaves, roots, shoots, etc. Endogenous promoters of genes related to herbicide tolerance and related phenotypes are particularly useful for driving expression of GAT nucleic acids, e.g., P450 monooxygenases, glutathione-S-transferases, homoglutathione-S-transferases, glyphosate oxidases and 5-enolpyruvylshikimate-2-phosphate synthases.

Tissue specific promoters can also be used to direct expression of heterologous structural genes, including the GAT polynucleotides described herein. Thus the promoters can be used in recombinant expression cassettes to drive expression of any

gene whose expression is desirable in the transgenic plants of the invention, e.g., GAT and/or other genes conferring herbicide resistance or tolerance, genes which influence other useful characteristics, e.g., heterosis. Similarly, enhancer elements, e.g., derived from the 5′ regulatory sequences or intron of a heterologous gene, can also be used to improve expression of a heterologous structural gene, such as a GAT polynucleotide.

In general, the particular promoter used in the expression cassette in plants depends on the intended application. Any of a number of promoters which direct transcription in plant cells can be suitable. The promoter can be either constitutive or inducible. In addition to the promoters noted above, promoters of bacterial origin which operate in plants include the octopine synthase promoter, the nopaline synthase promoter and other promoters derived from Ti plasmids. *See,* Herrera-Estrella et al. (1983) Nature 303:209. Viral promoters include the 35S and 19S RNA promoters of CaMV. *See*, Odell et al., (1985) Nature 313:810. Other plant promoters include the ribulose-1,3-bisphosphate carboxylase small subunit promoter and the phaseolin promoter. The promoter sequence from the E8 gene (*see*, Deikman and Fischer (1988) EMBO J 7:3315) and other genes are also favorably used. Promoters specific for monocotyledonous species are also considered (McElroy D., Brettell R.I.S. 1994. Foreign gene expression in transgenic cereals. Trends Biotech., 12:62-68.) Alternatively, novel promoters with useful characteristics can be identified from any viral, bacterial, or plant source by methods, including sequence analysis, enhancer or promoter trapping, and the like, known in the art.

In preparing expression vectors of the invention, sequences other than the promoter and the GAT encoding gene are also favorably used. If proper polypeptide expression is desired, a polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. Signal/localization peptides, which, e.g., facilitate translocation of the expressed polypeptide to internal organelles (e.g., chloroplasts) or extracellular secretion, can also be employed.

The vector comprising the GAT polynucleotide also can include a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide tolerance, particularly antibiotic tolerance, such as tolerance to kanamycin, G418, bleomycin, hygromycin, or herbicide tolerance, such as tolerance to chlorosulfuron, or phophinothricin. Reporter genes, which are used to monitor gene expression and protein localization via visualizable reaction products (e.g., beta-glucuronidase, beta-

galactosidase, and chloramphenicol acetyltransferase) or by direct visualization of the gene product itself (e.g., green fluorescent protein, GFP; Sheen et al. (1995) The Plant Journal 8:777) can be used for, e.g., monitoring transient gene expression in plant cells. Transient expression systems can be employed in plant cells, for example, in screening plant cell cultures for herbicide tolerance activities.

PLANT TRANSFORMATION

Protoplasts

Numerous protocols for establishment of transformable protoplasts from a variety of plant types and subsequent transformation of the cultured protoplasts are available in the art and are incorporated herein by reference. For examples, *see*, Hashimoto et al. (1990) Plant Physiol. 93:857; Fowke and Constabel (eds)(1994) Plant Protoplasts; Saunders et al. (1993) Applications of Plant In Vitro Technology Symposium, UPM 16-18; and Lyznik et al. (1991) BioTechniques 10:295, each of which is incorporated herein by reference.

Chloroplasts

Chloroplasts are a site of action of some herbicide tolerance activities, and, in some instances, the GAT polynucleotide is fused to a chloroplast transit sequence peptide to facilitate translocation of the gene products into the chloroplasts. In these cases, it can be advantageous to transform the GAT polynucleotide into the chloroplasts of the plant host cells. Numerous methods are available in the art to accomplish chloroplast transformation and expression (e.g., Daniell et al. (1998) Nature Biotechnology 16:346; O'Neill et al. (1993) The Plant Journal 3:729; Maliga (1993) TIBTECH 11:1). The expression construct comprises a transcriptional regulatory sequence functional in plants operably linked to a polynucleotide encoding the GAT polypeptide. Expression cassettes that are designed to function in chloroplasts (such as an expression cassette including a GAT polynucleotide) include the sequences necessary to ensure expression in chloroplasts. Typically, the coding sequence is flanked by two regions of homology to the chloroplastid genome to effect a homologous recombination with the chloroplast genome; often a selectable marker gene is also present within the flanking plastid DNA sequences to facilitate selection of genetically stable transformed chloroplasts in the resultant transplastonic plant cells (*see*, e.g., Maliga (1993) and Daniell (1998), and references cited therein).

General transformation methods

DNA constructs of the invention can be introduced into the genome of the desired plant host by a variety of conventional techniques. Techniques for tranforming a wide variety of higher plant species are well known and described in the technical and scientific literature. *See*, e.g., Payne, Gamborg, Croy, Jones, etc. all *supra*, as well as, e.g., Weising et al. (1988) Ann. Rev. Genet. 22:421.

For example, DNAs can be introduced directly into the genomic DNA of a plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs can be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the Agrobacterium host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the plant cell is infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al (1984) EMBO J 3:2717. Electroporation techniques are described in Fromm et al. (1985) Proc Nat'l Acad Sci USA 82:5824. Ballistic transformation techniques are described in Klein et al. (1987) Nature 327:70; and Weeks et al. Plant Physiol 102:1077.

In some embodiments, Agrobacterium mediated transformation techniques are used to transfer the GAT sequences of the invention to transgenic plants. Agrobacterium-mediated transformation is widely used for the transformation of dicots, however, certain monocots can also be transformed by Agrobacterium. For example, Agrobacterium transformation of rice is described by Hiei et al. (1994) Plant J. 6:271; US Patent No. 5,187,073; US Patent No. 5,591,616; Li et al. (1991) Science in China34:54; and Raineri et al. (1990) Bio/Technology 8:33. Transformed maize, barley, triticale and asparagus by Agrobacterium mediated transformation have also been described (Xu et al. (1990) Chinese J Bot 2:81).

Agrobacterium mediated transformation techniques take advantage of the ability of the tumor-inducing (Ti) plasmid of *A. tumefaciens* to integrate into a plant cell genome, to co-transfer a nucleic acid of interest into a plant cell. Typically, an expression vector is produced wherein the nucleic acid of interest, such as a GAT polynucleotide of the invention, is ligated into an autonomously replicating plasmid which also contains T-

DNA sequences. T-DNA sequences typically flank the expression casssette nucleic acid of interest and comprise the integration sequences of the plasmid. In addition to the expression cassette, T-DNA also typically include a marker sequence, e.g., antibiotic resistance genes. The plasmid with the T-DNA and the expression cassette are then transfected into Agrobacterium cells. Typically, for effective tranformation of plant cells, the *A. tumefaciens* bacterium also possesses the necessary *vir* regions on a plasmid, or integrated into its chromosome. For a discussion of Agrobacterium mediated transformation, *see*, Firoozabady and Kuehnle, (1995) Plant Cell Tissue and Organ Culture Fundamental Methods, Gamborg and Phillips (eds.).

Regeneration of Transgenic Plants

Transformed plant cells which are derived by plant transformation techniques, including those discussed above, can be cultured to regenerate a whole plant which possesses the transformed genotype (i.e., a GAT polynucleotide), and thus the desired phenotype, such as acquired resistance (i.e., tolerance) to glyphosate or a glyphosate analog. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Alternatively, selection for glyphosate resistance conferred by the GAT polynucleotide of the invention can be performed. Plant regeneration from cultured protoplasts is described in Evans et al. (1983) Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp 124-176, Macmillan Publishing Company, New York; and Binding (1985) Regeneration of Plants, Plant Protoplasts pp 21-73, CRC Press, Boca Raton. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. (1987) Ann Rev of Plant Phys 38:467. *See also*, e.g., Payne and Gamborg. After transformation with Agrobacterium, the explants typically are transferred to selection medium. One of skill will realize that the selection medium depends on the selectable marker that was co-transfected into the explants. After a suitable length of time, transformants will begin to form shoots. After the shoots are about 1-2 cm in length, the shoots should be transferred to a suitable root and shoot medium. Selection pressure should be maintained in the root and shoot medium.

Typically, the transformants will develop roots in about 1-2 weeks and form plantlets. After the plantlets are about 3-5 cm in height, they are placed in sterile soil in fiber pots. Those of skill in the art will realize that different acclimation procedures are

used to obtain transformed plants of different species. For example, after developing a root and shoot, cuttings, as well as somatic embryos of transformed plants, are transferred to medium for establishment of plantlets. For a description of selection and regeneration of transformed plants, *see*, e.g., Dodds and Roberts (1995) Experiments in Plant Tissue Culture, 3rd Ed., Cambridge University Press.

There are also methods for Agrobacterium transformation of Arabidopsis using vacuum infiltration (Bechtold N., Ellis J. and Pelletier G,, 1993, In planta Agrobacterium mediated gene transfer by infiltration of adult Arabidopsis thaliana plants. CR Acad Sci Paris Life Sci 316:1194-1199) and simple dipping of flowering plants (Desfeux, C., Clough S.J., and Bent A.F., 2000, Female reproductive tissues are the primary target of Agrobacterium-mediated transformation by the Arabidopsis floral-dip method. Plant Physiol. 123:895-904). Using these methods, transgenic seed are produced without the need for tissue culture.

There are plant varieties for which effective Agrobacterium-mediated transformation protocols have yet to be developed. For example, successful tissue transformation coupled with regeneration of the transformed tissue to produce a transgenic plant has not been reported for some of the most commercially relevant cotton cultivars. Nevertheless, an approach that can be used with these plants involves stably introducing the polynucleotide into a related plant variety via Agrobacterium-mediated transformation, confirming operability, and then transferring the transgene to the desired commercial strain using standard sexual crossing or back-crossing techniques. For example, in the case of cotton, Agrobacterium can be used to transform a Coker line of *Gossypium hirustum* (e.g., Coker lines 310, 312, 5110 Deltapine 61 or Stoneville 213), and then the transgene can be introduced into another more commercially relevant *G. hirustum* cultivar by back-crossing.

The transgenic plants of this invention can be characterized either genotypically or phenotypically to determine the presence of the GAT polynucleotide of the invention. Genotypic analysis can be performed by any of a number of well-known techniques, including PCR amplification of genomic DNA and hybridization of genomic DNA with specific labeled probes. Phenotypic analysis includes, e.g., survival of plants or plant tissues exposed to a selected herbicide such as glyphosate.

Essentially any plant can be transformed with the GAT polynucleotides of the invention. Suitable plants for the transformation and expression of the novel GAT polynucleotides of this invention include agronomically and horticulturally important

species. Such species include, but are not restricted to members of the families: Graminae (including corn, rye, triticale, barley, millet, rice, wheat, oats, etc.); Leguminosae (including pea, beans, lentil, peanut, yam bean, cowpeas, velvet beans, soybean, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, and sweetpea); Compositae (the largest family of vascular plants, including at least 1,000 genera, including important commercial crops such as sunflower) and Rosaciae (including raspberry, apricot, almond, peach, rose, etc.), as well as nut plants (including, walnut, pecan, hazelnut, etc.), and forest trees (including *Pinus, Quercus, Pseutotsuga, Sequoia, Populus,*etc.)

Additional targets for modification by the GAT polynucleotides of the invention, as well as those specified above, include plants from the genera: *Agrostis, Allium, Antirrhinum, Apium, Arachis, Asparagus, Atropa, Avena* (e.g., oats), *Bambusa, Brassica, Bromus, Browaalia, Camellia, Cannabis, Capsicum, Cicer, Chenopodium, Chichorium, Citrus, Coffea, Coix, Cucumis, Curcubita, Cynodon, Dactylis, Datura, Daucus, Digitalis, Dioscorea, Elaeis, Eleusine, Festuca, Fragaria, Geranium, Gossypium, Glycine, Helianthus, Heterocallis, Hevea, Hordeum* (e.g., barley), *Hyoscyamus, Ipomoea, Lactuca, Lens, Lilium, Linum, Lolium, Lotus, Lycopersicon, Majorana, Malus, Mangifera, Manihot, Medicago, Nemesia, Nicotiana, Onobrychis, Oryza* (e.g., rice), *Panicum, Pelargonium, Pennisetum* (e.g., millet), *Petunia, Pisum, Phaseolus, Phleum, Poa, Prunus, Ranunculus, Raphanus, Ribes, Ricinus, Rubus, Saccharum, Salpiglossis, Secale* (e.g., rye), *Senecio, Setaria, Sinapis, Solanum, Sorghum, Stenotaphrum, Theobroma, Trifolium, Trigonella, Triticum* (e.g., wheat), *Vicia, Vigna, Vitis, Zea* (e.g., corn), and the *Olyreae*, the *Pharoideae* and many others. As noted, plants in the family *Graminae* are a particularly target plants for the methods of the invention.

Common crop plants which are targets of the present invention include corn, rice, triticale, rye, cotton, soybean, sorghum, wheat, oats, barley, millet, sunflower, canola, peas, beans, lentils, peanuts, yam beans, cowpeas, velvet beans, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, sweetpea and nut plants (e.g., walnut, pecan, etc).

In one aspect, the invention provides a method for producing a crop by growing a crop plant that is glyphosate-tolerant as a result of being transformed with a gene encoding a glyphosate N-acteyltransferase, under conditions such that the crop plant produces a crop, and harvesting the crop. Preferably, glyphosate is applied to the plant, or in the vicinity of the plant, at a concentration effective to control weeds without preventing the transgenic crop plant from growing and producing the crop. The application of

glyphosate can be before planting, or at any time after planting up to and including the time of harvest. Glyphosate can be applied once or multiple times. The timing of glyphosate application, amount applied, mode of application, and other parameters will vary based upon the specific nature of the crop plant and the growing environment, and can be readily determined by one of skill in the art. The invention further provides the crop produced by this method.

The invention provides for the propagation of a plant containing a GAT polynucleotide transgene. The plant can be, for example, a monocot or a dicot. In one aspect, propagation entails crossing a plant containing a GAT polynucleotide transgene with a second plant, such that at least some progeny of the cross display glyphosate tolerance.

In one aspect, the invention provides a method for selectively controlling weeds in a field where a crop is being grown. The method involves planting crop seeds or plants that are glyphosate-tolerant as a result of being transformed with a gene encoding a GAT, e.g., a GAT polynucleotide, and applying to the crop and any weeds a sufficient amount of glyphosate to control the weeds without a significant adverse impact on the crops. It is important to note that it is not necessary for the crop to be totally insensitive to the herbicide, so long as the benefit derived from the inhibition of weeds outweighs any negative impact of the glyphosate or glyphosate analog on the crop or crop plant.

In another aspect, the invention provides for use of a GAT polynucleotide as a selectable marker gene. In this embodiment of the invention, the presence of the GAT polynucleotide in a cell or organism confers upon the cell or organism the detectable phenotypic trait of glyphosate resistance, thereby allowing one to select for cells or organisms that have been transformed with a gene of interest linked to the GAT polynucleotide. Thus, for example, the GAT polynucleotide can be introduced into a nucleic acid construct, e.g., a vector, thereby allowing for the identification of a host (e.g., a cell or transgenic plant) containing the nucleic acid construct by growing the host in the presence of glyphosate and selecting for the ability to survive and/or grow at a rate that is discernibly greater than a host lacking the nucleic acid construct would survive or grow. A GAT polynucleotide can be used as a selectable marker in a wide variety of hosts that are sensitive to glyphosate, including plants, most bacteria (including *E. coli*), actinomycetes, yeasts, algae and fungi. One benefit of using herbicide resistance as a marker in plants, as opposed to conventional antibiotic resistance, is that it obviates the concern of some members of the public that antibiotic resistance might escpe into the

environment. Some experimental data from experiments demonstrating the use of a GAT polynucleotide as a selectable marker in diverse host systems are described in the Examples section of this specification.

Selection of gat polynucleotides conferring enhanced glyphosate resistance in transgenic plants.

Libraries of GAT encoding nucleic acids diversified according to the methods described herein can be selected for the ability to confer resistance to glyphosate in transgenic plants. Following one or more cycles of diversification and selection, the modified GAT genes can be used as a selection marker to facilitate the production and evaluation of transgenic plants and as a means of conferring herbicide resistance in experimental or agricultural plants. For example, after diversification of any one or more of SEQ ID NO:1 to SEQ ID NO:5 to produce a library of diversified GAT polynucleotides, an initial functional evaluation can be performed by expressing the library of GAT encoding sequences in *E. coli.* The expressed GAT polypeptides can be purified, or partially purified as described above, and screened for improved kinetics by mass spectrometry. Following one or more preliminary rounds of diversification and selection, the polynucleotides encoding improved GAT polypeptides are cloned into a plant expression vector, operably linked to, e.g., a strong constitutive promoter, such as the CaMV 35S promoter. The expression vectors comprising the modified GAT nucleic acids are transformed, typically by Agrobacterium mediated transformation, into *Arabidopsis thaliana* host plants. For example, *Arabidopsis* hosts are readily transformed by dipping inflorescences into solutions of Agrobacterium and allowing them to grow and set seed. Thousands of seeds are recovered in approximately 6 weeks. The seeds are then collected in bulk from the dipped plants and germinated in soil. In this manner it is possible to generate several thousand independently transformed plants for evaluation, constituting a high throughput (HTP) plant transformation format. Bulk grown seedlings are sprayed with glyphosate and surviving seedlings exhibiting glyphosate resistance survive the selection process, whereas non-transgenic plants and plants incorporating less favorable modified GAT nucleic acids are damaged or killed by the herbicide treatment. Optionally, the GAT encoding nucleic acids conferring improved resistance to glyphosate are recovered, e.g., by PCR amplification using T-DNA primers flanking the library inserts, and used in further diversification procedures or to produce additional transgenic plants of the same or different species. If desired, additional rounds of diversification and selection

can be performed using increasing concentrations of glyphosate in each subsequent selection. In this manner, GAT polynucleotides and polypeptides conferring resistance to concentrations of glyphosate useful in field conditions can be obtained.

Herbicide Resistance

The mechanism of glyphosate resistance of the present invention can be combined with other modes of glyphosate resistance known in the art to produce plants and plant explants with superior glyphosate resistance. For example, glyphosate-tolerant plants can be produced by inserting into the genome of the plant the capacity to produce a higher level of 5-enolpyruvylshikimate-3-phosphate synthase (EPSP) as more fully described in U.S. Patent Nos. 6,248,876 B1; 5,627,061; 5,804,425; 5,633,435; 5,145,783; 4,971,908; 5,312,910; 5,188,642; 4,940,835; 5,866,775; 6,225,114 B1; 6,130,366; 5,310,667; 4,535,060; 4,769,061; 5,633,448; 5,510,471; Re. 36,449; RE 37,287 E; and 5,491,288; and international publications WO 97/04103; WO 00/66746; WO 01/66704; and WO 00/66747, which are incorporated herein by reference in their entireties for all purposes. Glyphosate resistance is also imparted to plants that express a gene that encodes a glyphosate oxido-reductase enzyme as described more fully in U.S. Patent Nos. 5,776,760 and 5,463,175, which are incorporated herein by reference in their entireties for all purposes.

Further, the mechanism of glyphosate resistance of the present invention may be combined with other modes of herbicide resistance to provide plants and plant explants that are resistant to glyphosate and one or more other herbicides. For example, the hydroxyphenylpyruvatedioxygenases are enzymes that catalyze the reaction in which para-hydroxyphenylpyruvate (HPP) is transformed into homogentisate. Molecules which inhibit this enzyme, and which bind to the enzyme in order to inhibit transformation of the HPP into homogentisate are useful as herbicides. Plants more resistant to certain herbicides are described in U.S Patent Nos. 6,245,968 B1; 6,268,549; and 6,069,115; and international publication WO 99/23886, which are incorporated herein by reference in their entireties for all purposes.

Sulfonylurea and imidazolinone herbicides also inhibit growth of higher plants by blocking acetolactate synthase (ALS) or acetohydroxy acid synthase (AHAS). The production of sulfonylurea and imidazolinone tolerant plants is described more fully in U.S Patent Nos. 5,605,011; 5,013,659; 5,141,870; 5,767,361; 5,731,180; 5,304,732; 4,761,373; 5,331,107; 5,928,937; and 5,378,824; and international publication WO 96/33270, which are incorporated herein by reference in their entireties for all purposes.

Glutamine synthetase (GS) appears to be an essential enzyme necessary for the development and life of most plant cells. Inhibitors of GS are toxic to plant cells. Glufosinate herbicides have been developed based on the toxic effect due to the inhibition of GS in plants. These herbicides are non-selective. They inhibit growth of all the different species of plants present, causing their total destruction. The development of plants containing an exogenous phosphinothricin acetyl transferase is described in U.S. Patent Nos. 5,969,213; 5,489,520; 5,550,318; 5,874,265; 5,919,675; 5,561,236; 5,648,477; 5,646,024; 6,177,616 B1; and 5,879,903, which are incorporated herein by reference in their entireties for all purposes.

Protoporphyrinogen oxidase (protox) is necessary for the production of chlorophyll, which is necessary for all plant survival. The protox enzyme serves as the target for a variety of herbicidal compounds. These herbicides also inhibit growth of all the different species of plants present, causing their total destruction. The development of plants containing altered protox activity which are resistant to these herbicides are described in U.S. Patent Nos. 6,288,306 B1; 6,282,837 B1; and 5,767,373; and international publication WO 01/12825, which are incorporated herein by reference in their entireties for all purposes.

EXAMPLES

The following examples are illustrative and not limiting. One of skill will recognize a variety of non-critical parameters that can be altered to achieve essentially similar results.

EXAMPLE 1: ISOLATING NOVEL NATIVE GAT POLYNUCLEOTIDES

Five native GAT polynucleotides (i.e., GAT polynucleotides that occur naturally in a non-genetically modified organism) were discovered by expression cloning of sequences from *Bacillus* strains exhibiting GAT activity. Their nucleotide sequences were determined and are provided herein as SEQ ID NO:1 to SEQ ID NO:5. Briefly, a collection of approximately 500 *Bacillus* and *Pseudomonas* strains were screened for native ability to N-acetylate glyphosate. Strains were grown in LB overnight, harvested by centrifugation, permeabilizied in dilute toluene, and then washed and resuspended in a reaction mix containing buffer, 5 mM glyphosate, and 200 μM acetyl-CoA. The cells were incubated in the reaction mix for between 1 and 48 hours, at which time an equal volume of methanol was added to the reaction. The cells were then pelleted by centrifugation and the supernatant was filtered before analysis by parent ion mode mass

spectrometry. The product of the reaction was positively identified as N-acetylglyphosate by comparing the mass spectrometry profile of the reaction mix to an N-acetylglyphosate standard as shown in Figure 2. Product detection was dependent on inclusion of both substrates (acetylCoA and glyphosate) and was abolished by heat denaturing the bacterial cells.

Individual GAT polynucleotides were then cloned from the identified strains by functional screening. Genomic DNA was prepared and partially digested with Sau3A1 enzyme. Fragments of approximately 4 Kb were cloned into an E. coli expression vector and transformed into electrocompetent E. coli. Individual clones exhibiting GAT activity were identified by mass spectrometry following a reaction as described previously except that the toluene wash was replaced by permeabilization with PMBS. Genomic fragments were sequenced and the putative GAT polypeptide-encoding open reading frame identified. Identity of the GAT gene was confirmed by expression of the open reading frame in E. coli and detection of high levels of N-acetylglyphosate produced from reaction mixtures.

EXAMPLE 2: CHARACTERIZATION OF A GAT POLYPEPTIDE ISOLATED FROM *B.LICHENIFORMIS* STRAIN B6.

Genomic DNA from *B. licheniformis* strain B6 was purified, partially digested with Sau3A1 and fragments of 1-10 Kb were cloned into an E. coli expression vector. A clone with a 2.5 kb insert conferred the glyphosate N-acetyltransferase (GAT) activity on the E. coli host as determined with mass spectrometry analysis. Sequencing of the insert revealed a single complete open reading frame of 441 base pairs. Subsequent cloning of this open reading frame confirmed that it encoded the GAT enzyme. A plasmid, pMAXY2120, shown in figure 4, with the gene encoding the GAT enzyme of B6 was transformed into *E. coli* strain XL1 Blue. A 10% innoculum of a saturated culture was added to Luria broth, and the culture was incubated at 37° C for 1 hr. Expression of GAT was induced by the addition of IPTG at a concentration of 1 mM. The culture was incubated a further 4 hrs, following which, cells were harvested by centrifugation and the cell pellet stored at –80° C.

Lysis of the cells was effected by the addition of 1 ml of the following buffer to 0.2 g of cells: 25 mM HEPES, pH 7.3, 100 mM KCl and 10% methanol (HKM) plus 0.1 mM EDTA, 1 mM DTT, 1 mg/ml chicken egg lysozyme, and a protease inhibitor

cocktail obtained from Sigma and used according to the manufacturer's recommendations. After 20 minutes incubation at room temperature (e.g., 22-25° C), lysis was completed with brief sonication. The lysate was centrifuged and the supernatant was desalted by passage through Sephadex G25 equilibrated with HKM. Partial purification was obtained by affinity chromatography on CoA Agarose (Sigma). The column was equilibrated with HKM and the clarified extract allowed to pass through under hydrostatic pressure. Non-binding proteins were removed by washing the column with HKM, and GAT was eluted with HKM containing 1 mM Coenzyme A. This procedure provided 4-fold purification. At this stage, approximately 65% of the protein staining observed on an SDS polyacrylamide gel loaded with crude lysate was due to GAT, with another 20% due to chloramphenicol acetyltransferase encoded by the vector.

Purification to homogeneity was obtained by gel filtration of the partially purified protein through Superdex 75 (Pharmacia). The mobile phase was HKM, in which GAT activity eluted at a volume corresponding to a molecular radius of 17 kD. This material was homogeneous as judged by Coomassie staining of a 3 μg sample of GAT subjected to SDS polyacrylamide gel electrophoresis on a 12% acrylamide gel, 1 mm thickness. Purification was achieved with a 6-fold increase in specific activity.

The apparent K_M for glyphosate was determined on reaction mixtures containing saturating (200 μM) Acetyl CoA, varying concentrations of glyphosate, and 1 μM purified GAT in buffer containing 5 mM morpholine adjusted to pH 7.7 with acetic acid and 20 % ethylene glycol. Initial reaction rates were determined by continuous monitoring of the hydrolysis of the thioester bond of Acetyl CoA at 235 nm (E = 3.4 OD/mM/cm). Hyperbolic saturation kinetics were observed (Figure 5), from which an apparent K_M of 2.9 ± 0.2 (SD) mM was obtained.

The apparent K_M for AcCoA was determined on reaction mixtures containing 5 mM glyphosate, varying concentrations of Acetyl CoA, and 0.19 μM GAT in buffer containing 5 mM morpholine adjusted to pH 7.7 with acetic acid and 50% methanol. Initial reaction rates were determined using mass spectrometric detection of N-acetyl glyphosate. Five μl were repeatedly injected to the instrument and reaction rates were obtained by plotting reaction time vs area of the integrated peak (Figure 6). Hyperbolic saturation kinetics were observed (Figure 7), from which an apparent K_M of 2 μM was derived. From values for Vmax obtained at a known concentration of enzyme, a kcat of 6/min was calculated.

EXAMPLE 3: MASS SPECTROMETRY (MS) SCREENING PROCESS

Sample (5 ul) is drawn from a 96-well microtiter plate at a speed of one sample every 26 seconds and injected into the mass spectrometer (Micromass Quattro LC, triple quadrupole mass spectrometer) without any separation. The sample is carried into the mass spectrometer by a mobile phase of water/methanol (50:50) at a flow rate of 500 Ul/min. Each injected sample is ionized by negative electrospray ionization process (needle voltage, –3.5 KV; cone voltage, 20 V; source temperature, 120 C; desolvation temperature, 250 C; cone gas flow, 90 L/Hr; and desolvation gas flow, 600 L/Hr). The molecular ions (m/z 210) formed during this process arre selected by the first quadrupole for performing collison induced dissociation (CID) in the second quadrupole, where the pressure is set at 5 x 10^{-4} mBar and the collision energy is adjusted to 20 Ev. The third quadrupole is set for only allowing one of the daughter ions (m/z 124) produced from the parent ions (m/z 210) to get into the detector for signal recording. The first and third quadupoles are set at unit resolution, while the photomultiplier is operated at 650 V. Pure N-acetylglyphosate standards are used for comparison and peak integration used to estimate concentrations. It is possible to detect less than 200 Nm N-acetylglyphosate by this method.

EXAMPLE 4: DETECTION OF NATIVE OR LOW ACTIVITY GAT ENZYMES

Native or low activity GAT enzymes typically have Kcat of approximately 1 min^{-1} and K_M for glyphosate of 1.5-10 Mm. K_M for acetylCoA is typically less than 25 μM.

Bacterial cultures are grown in rich medium in deep 96-well plates and 0.5 ml stationary phase cells are harvested by centrifugation, washed with 5 mM morpholine acetate pH 8, and resuspended in 0.1 ml reaction mix containing 200 μM ammonium acetylCoA, 5 mM ammonium glyphosate, and 5 μg/ml PMBS (Sigma) in 5 mM morpholine acetate, pH 8. The PMBS permeabilizes the cell membrane allowing the substrates and products to move from the cells to the buffer without releasing the entire cellular contents. Reactions are carried out at 25-37°C for 1-48 hours. The reactions are quenched with an equal volume of 100% ethanol and the entire mixture is filtered on a 0.45 μm MAHV Multiscreen filter plate (Millipore). Samples are analyzed using a mass spectrometer as desribed above and compared to synthetic N-acetylglyphosate standards.

EXAMPLE 5: DETECTION OF HIGH ACTIVITY GAT ENZYMES

High activity GAT enzymes typically have kcat up to 400 min^{-1} and K_M below 0.1 mM glyphosate.

Genes coding for GAT enzymes are cloned into *E. coli* expression vectors such as pQE80 (Qiagen) and introduced into *E. coli* strains such as XL1 Blue (Stratagene). Cultures are grown in 150 ul rich medium (such as LB with 50 ug/ml carbenicllin) in shallow U-bottom 96-well polystyrene plates to late-log phase and diluted 1:9 with fresh medium containing 1 mM IPTG (USB). After 4-8 hours induction, cells are harvested, washed with 5mM morpholine acetate pH 6.8 and resuspended in an equal volume of the same morpholine buffer. Reactions are carried out with up to 10 ul of washed cells. At higher activity levels, the cells are first diluted up to 1:200 and 5 ul is added to100 ul reaction mix. To measure GAT activity, the same reaction mix as described for low activity can be used. However, for detecting highly active GAT enzymes the glyphosate concentration is reduced to 0.15 – 0.5 mM, the pH is reduced to 6.8, and reactions are carried out for 1 hour at 37°C. Reaction workup and MS detection are as described herein.

EXAMPLE 6: PURIFICATION OF GAT ENZYMES

Enzyme purification is achieved by affinity chromatography of cell lysates on CoA-agarose and gel-filtration on Superdex-75. Quantities of purified GAT enzyme up to 10 mg are obtained as follows: A 100-ml culture of *E. coli* carrying a GAT polynucleotide on a pQE80 vector and grown overnight in LB containing 50 ug/ml carbenicillin is used to inoculate 1 L of LB plus 50 ug/ml carbenicillin. After 1 hr, IPTG is added to 1 mM, and the culture is grown a further 6 hr. Cells are harvested by centrifugation. Lysis is effected by suspending the cells in 25 mM HEPES (pH 7.2), 100 mM KCl, 10% methanol (termed HKM), 0.1 mM EDTA, 1 mM DTT, protease inhibitor cocktail supplied by Sigma-Aldrich and 1 mg/ml of chicken egg lysozyme. After 30 minutes at room temperature, the cells are briefly sonicated. Particulate material is removed by centrifugation, and the lysate is passed through a bed of coenzyme A-Agarose. The column is washed with several bed volumes of HKM and GAT is eluted in 1.5 bed volumes of HKM containing 1 mM acetyl-coenzyme A. GAT in the eluate is concentrated by its retention above a Centricon YM 50 ultrafiltration membrane. Further purification is obtained by passing the protein through a Superdex 75 column through a series of 0.6-ml injections. The peak of GAT activity elutes at a volume corresponding to

a molecular weight of 17 kD. This method results in purification of GAT enzyme to homogeneity with >85% recovery. A similar procedure is used to obtain 0.1 to 0.4 mg quantities of up to 96 shuffled variants at a time. The volume of induced culture is reduced to 1 to 10 ml, coenzyme A-Agarose affinity chromatography is performed in 0.15-ml columns packed in an MAHV filter plate (Millipore) and Superdex 75 chromatography is omitted.

EXAMPLE 7: STANDARD PROTOCOL FOR DETERMINATION OF K_{CAT} AND K_M

K_{cat} and K_M for glyphosate of purified protein are determined using a continuous spectrophotometric assay, in which hydrolysis of the sulfoester bond of AcCoA is monitored at 235 nm. Reactions are performed at ambient temperature (about 23°C) in the wells of a 96-well assay plate, with the following components present in a final volume of 0.3 ml: 20 mM HEPES, pH 6.8, 10% ethylene glycol, 0.2 mM acetyl coenzyme A, and various concentration of ammonium glyphosate. In comparing the kinetics of two GAT enzymes, both enzymes should be assayed under the same condition, e.g., both at 23°C. K_{cat} is calculated from V_{max} and the enzyme concentration, determined by Bradford assay. K_M is calculated from the initial reaction rates obtained from concentrations of glyphosate ranging from 0.125 to 10 mM, using the Lineweaver-Burke transformation of the Michaelis-Menten equation. K_{cat}/K_M is determined by dividing the value determined for K_{cat} by the value determined for K_M.

Using this methodology, kinetic parameters for a number of GAT polypeptides exemplified herein have been determined. For example, the K_{cat}, K_M and K_{cat}/K_M for the GAT polypeptide corresponding to SEQ ID NO:445 have been determined to be 322 min^{-1}, 0.5 mM and 660 $mM^{-1}min^{-1}$, respectively, using the assay conditions described above. The K_{cat}, K_M and K_{cat}/K_M for the GAT polypeptide corresponding to SEQ ID NO:457 have been determined to be 118 min^{-1}, 0.1 mM and 1184 $mM^{-1}min^{-1}$, respectively, using the assay conditions described above. The K_{cat}, K_M and K_{cat}/K_M for the GAT polypeptide corresponding to SEQ ID NO:300 have been determined to be 296 min^{-1}, 0.65 mM and 456 $mM^{-1}min^{-1}$, respectively, using the assay conditions described above. One of skill in the art can use these numbers to confirm that a GAT activity assay is generating kinetic parameters for a GAT suitable for comparison with the values given herein. For example, the conditions used to compare the activity of GATs should yield the same kinetic constants for SEQ ID NOS: 300, 445 and 457 (within normal experimental

variance) as those reported herein, if the conditions are going to be used to compare a test GAT with the GAT polypeptides exemplified herein. Kinetic parameters for a number of GAT polypeptide variants were determined according to this methodology and are provided in Tables 3, 4 and 5.

Table 3. GAT polypeptide k_{cat} values

SEQ ID NO.	Clone ID	$K_{cat}(min^{-1})$
SEQ ID NO:263	13_10F6	48.6
SEQ ID NO:264	13_12G6	52.1
SEQ ID NO:265	14_2A5	280.8
SEQ ID NO:266	14_2C1	133.4
SEQ ID NO:267	14_2F11	136.9
SEQ ID NO:268	CHIMERA	155.4
SEQ ID NO:269	10_12D7	77.3
SEQ ID NO:270	10_15F4	37.6
SEQ ID NO:271	10_17D1	176.2
SEQ ID NO:272	10_17F6	47.9
SEQ ID NO:273	10_18G9	24
SEQ ID NO:274	10_1H3	76.2
SEQ ID NO:275	10_20D10	86.2
SEQ ID NO:276	10_23F2	101.3
SEQ ID NO:277	10_2B8	108.4
SEQ ID NO:278	10_2C7	135
SEQ ID NO:279	10_3G5	87.4
SEQ ID NO:280	10_4H7	112
SEQ ID NO:281	10_6D11	62.4
SEQ ID NO:282	10_8C6	21.7
SEQ ID NO:283	11C3	2.8
SEQ ID NO:284	11G3	15.6
SEQ ID NO:285	11H3	1.2
SEQ ID NO:286	12_1F9	80.4
SEQ ID NO:287	12_2G9	151.4
SEQ ID NO:288	12_3F1	44.1
SEQ ID NO:289	12_5C10	89.6
SEQ ID NO:290	12_6A10	54.7
SEQ ID NO:291	12_6D1	49
SEQ ID NO:292	12_6F9	89.1
SEQ ID NO:293	12_6H6	90.5
SEQ ID NO:294	12_7D6	53.9
SEQ ID NO:295	12_7G11	234.5
SEQ ID NO:296	12F5	3.1
SEQ ID NO:297	12G7	2.3
SEQ ID NO:298	1_2H6	9.3
SEQ ID NO:299	13_12G12	36.1
SEQ ID NO:300	13_6D10	296.5
SEQ ID NO:301	13_7A7	117
SEQ ID NO:302	13_7B12	68.9
SEQ ID NO:303	13_7C1	48.1
SEQ ID NO:304	13_8G6	33.7
SEQ ID NO:305	13_9F6	59
SEQ ID NO:306	14_10C9	127
SEQ ID NO:307	14_10H3	105.2
SEQ ID NO:308	14_10H9	127.2

SEQ ID NO:309	14_11C2	108.7
SEQ ID NO:310	14_12D8	62.1
SEQ ID NO:311	14_12H6	91.1
SEQ ID NO:312	14_2B6	34.2
SEQ ID NO:313	14_2G11	69.4
SEQ ID NO:314	14_3B2	68.7
SEQ ID NO:315	14_4H8	198.8
SEQ ID NO:316	14_6A8	43.7
SEQ ID NO:317	14_6B10	134.7
SEQ ID NO:318	14_6D4	256
SEQ ID NO:319	14_7A11	197.2
SEQ ID NO:320	14_7A1	155.8
SEQ ID NO:321	14_7A9	245.9
SEQ ID NO:322	14_7G1	136.7
SEQ ID NO:323	14_7H9	64.4
SEQ ID NO:324	14_8F7	90.5
SEQ ID NO:325	15_10C2	69.9
SEQ ID NO:326	15_10D6	67.1
SEQ ID NO:327	15_11F9	76.4
SEQ ID NO:328	15_11H3	61.9
SEQ ID NO:329	15_12A8	77.1
SEQ ID NO:330	15_12D6	148.6
SEQ ID NO:331	15_12D8	59.7
SEQ ID NO:332	15_12D9	59.7
SEQ ID NO:333	15_3F10	48.7
SEQ ID NO:334	15_3G11	71.5
SEQ ID NO:335	15_4F11	80.3
SEQ ID NO:336	15_4H3	93.3
SEQ ID NO:337	15_6D3	85.9
SEQ ID NO:338	15_6G11	36.9
SEQ ID NO:339	15_9F6	59.6
SEQ ID NO:340	15F5	0.5
SEQ ID NO:341	16A1	10.4
SEQ ID NO:342	16H3	3.5
SEQ ID NO:343	17C12	3.2
SEQ ID NO:344	18D6	9.6
SEQ ID NO:345	19C6	2.2
SEQ ID NO:346	19D5	2.2
SEQ ID NO:347	20A12	2.8
SEQ ID NO:348	20F2	3.9
SEQ ID NO:349	2.10E+12	1.1
SEQ ID NO:350	23H11	7.1
SEQ ID NO:351	24C1	1.7
SEQ ID NO:352	24C6	2.7
SEQ ID NO:353	2.40E+08	8.9
SEQ ID NO:354	2_8C3	24.8
SEQ ID NO:355	2H3	16.1
SEQ ID NO:356	30G8	10.2
SEQ ID NO:357	3B_10C4	24.8
SEQ ID NO:358	3B_10G7	19.6
SEQ ID NO:359	3B_12B1	22.8
SEQ ID NO:360	3B_12D10	5.4
SEQ ID NO:361	3B_2E5	16.4
SEQ ID NO:362	3C_10H3	33.9
SEQ ID NO:363	3C_12H10	9.1
SEQ ID NO:364	3C_9H8	11.7
SEQ ID NO:365	4A_1B11	23.2
SEQ ID NO:366	4A_1C2	20.4

SEQ ID NO:367	4B_13E1	37.2
SEQ ID NO:368	4B_13G10	34.9
SEQ ID NO:369	4B_16E1	17
SEQ ID NO:370	4B_17A1	19.1
SEQ ID NO:371	4B_18F11	14.6
SEQ ID NO:372	4B_19C8	15.9
SEQ ID NO:373	4B_1G4	3.7
SEQ ID NO:374	4B_21C6	11.8
SEQ ID NO:375	4B_2H7	27
SEQ ID NO:376	4B_2H8	38.3
SEQ ID NO:377	4B_6D8	22.7
SEQ ID NO:378	4B_7E8	20.5
SEQ ID NO:379	4C_8C9	9
SEQ ID NO:380	4H1	1.3
SEQ ID NO:381	6_14D10	42.2
SEQ ID NO:382	6_15G7	48.4
SEQ ID NO:383	6_16A5	43.8
SEQ ID NO:384	6_16F5	35.2
SEQ ID NO:385	6_17C5	35.2
SEQ ID NO:386	6_18C7	32.2
SEQ ID NO:387	6_18D7	43
SEQ ID NO:388	6_19A10	86.8
SEQ ID NO:389	6_19B6	23.9
SEQ ID NO:390	6_19C3	23.1
SEQ ID NO:391	6_19C8	74.8
SEQ ID NO:392	6_20A7	40.4
SEQ ID NO:393	6_20A9	45.1
SEQ ID NO:394	6_20H5	19.5
SEQ ID NO:395	6_21F4	24.3
SEQ ID NO:396	6_22C9	47.4
SEQ ID NO:397	6_22D9	43.9
SEQ ID NO:398	6_22H9	17.4
SEQ ID NO:399	6_23H3	43.9
SEQ ID NO:400	6_23H7	46.2
SEQ ID NO:401	6_2H1	26.6
SEQ ID NO:402	6_3D6	41.7
SEQ ID NO:403	6_3G3	51.9
SEQ ID NO:404	6_3H2	57.2
SEQ ID NO:405	6_4A10	55
SEQ ID NO:406	6_4B1	27
SEQ ID NO:407	6_5D11	15.2
SEQ ID NO:408	6_5F11	40.1
SEQ ID NO:409	6_5G9	35.8
SEQ ID NO:410	6_6D5	55.3
SEQ ID NO:411	6_7D1	19.7
SEQ ID NO:412	6_8H3	44.7
SEQ ID NO:413	6_9G11	78.4
SEQ ID NO:414	6F1	10.1
SEQ ID NO:415	7_1C4	17.4
SEQ ID NO:416	7_2A10	14.5
SEQ ID NO:417	7_2A11	46.8
SEQ ID NO:418	7_2D7	54.9
SEQ ID NO:419	7_5C7	44.7
SEQ ID NO:420	7_9C9	65
SEQ ID NO:421	9_13F10	34.7
SEQ ID NO:422	9_13F1	31.6
SEQ ID NO:423	9_15D5	27.6
SEQ ID NO:424	9_15D8	107.3

SEQ ID NO:425	9_15H3	68.7
SEQ ID NO:426	9_18H2	25
SEQ ID NO:427	9_20F12	37.8
SEQ ID NO:428	9_21C8	28.6
SEQ ID NO:429	9_22B1	50.1
SEQ ID NO:430	9_23A10	21
SEQ ID NO:431	9_24F6	52.5
SEQ ID NO:432	9_4H10	101.3
SEQ ID NO:433	9_4H8	47.1
SEQ ID NO:434	9_8H1	74.8
SEQ ID NO:435	9_9H7	28
SEQ ID NO:436	9C6	13
SEQ ID NO:437	9H11	4
SEQ ID NO:438	0_4B10	190
SEQ ID NO:439	0_5B11	219
SEQ ID NO:440	0_5B3	143
SEQ ID NO:441	0_5B4	180
SEQ ID NO:442	0_5B8	143
SEQ ID NO:443	0_5C4	205
SEQ ID NO:444	0_5D11	224
SEQ ID NO:445	0_5D3	322
SEQ ID NO:446	0_5D7	244
SEQ ID NO:447	0_6B4	252
SEQ ID NO:448	0_6D10	111
SEQ ID NO:449	0_6D11	212
SEQ ID NO:450	0_6F2	175
SEQ ID NO:451	0_6H9	228
SEQ ID NO:452	10_4C10	69.6
SEQ ID NO:453	10_4D5	82.72
SEQ ID NO:454	10_4F2	231.04
SEQ ID NO:455	10_4F9	55.39
SEQ ID NO:456	10_4G5	176.65
SEQ ID NO:457	10_4H4	118.36
SEQ ID NO:458	11_3A11	55.66
SEQ ID NO:459	11_3B1	219.97
SEQ ID NO:460	11_3B5	194.61
SEQ ID NO:461	11_3C12	49.07
SEQ ID NO:462	11_3C3	214.02
SEQ ID NO:463	11_3C6	184.44
SEQ ID NO:464	11_3D6	55.3
SEQ ID NO:465	1_1G12	58.48
SEQ ID NO:466	1_1H1	291
SEQ ID NO:467	1_1H2	164
SEQ ID NO:468	1_1H5	94
SEQ ID NO:469	1_2A12	229
SEQ ID NO:470	1_2B6	138
SEQ ID NO:471	1_2C4	193
SEQ ID NO:472	1_2D2	124
SEQ ID NO:473	1_2D4	182
SEQ ID NO:474	1_2F8	161
SEQ ID NO:475	1_2H8	141
SEQ ID NO:476	1_3A2	181
SEQ ID NO:477	1_3D6	226
SEQ ID NO:478	1_3F3	167
SEQ ID NO:479	1_3H2	128
SEQ ID NO:480	1_4C5	254
SEQ ID NO:481	1_4D6	137
SEQ ID NO:482	1_4H1	236

SEQ ID NO:483	1_5H5	214
SEQ ID NO:484	1_6F12	209
SEQ ID NO:485	1_6H6	274
SEQ ID NO:486	3_11A10	135.41
SEQ ID NO:487	3_14F6	188.43
SEQ ID NO:488	3_15B2	104.13
SEQ ID NO:489	3_6A10	126.48
SEQ ID NO:490	3_6B1	263.08
SEQ ID NO:491	3_7F9	193.55
SEQ ID NO:492	3_8G11	99.14
SEQ ID NO:493	4_1B10	77.09
SEQ ID NO:494	5_2B3	56.75
SEQ ID NO:495	5_2D9	75.44
SEQ ID NO:496	5_2F10	54.72
SEQ ID NO:497	6_1A11	45.54
SEQ ID NO:498	6_1D5	42.92
SEQ ID NO:499	6_1F11	105.76
SEQ ID NO:500	6_1F1	69.81
SEQ ID NO:501	6_1H10	17.01
SEQ ID NO:502	6_1H4	85.91
SEQ ID NO:503	8_1F8	82.88
SEQ ID NO:504	8_1G2	67.47
SEQ ID NO:505	8_1G3	108.9
SEQ ID NO:506	8_1H7	101.24
SEQ ID NO:507	8_1H9	78.39
SEQ ID NO:508	GAT1_21F12	5.4
SEQ ID NO:509	GAT1_24G3	4.9
SEQ ID NO:510	GAT1_29G1	6.2
SEQ ID NO:511	GAT1_32G1	4.5
SEQ ID NO:512	GAT2_15G8	4.5
SEQ ID NO:513	GAT2_19H8	4.1
SEQ ID NO:514	GAT2_21F1	4.2

Table 4. GAT polypeptide (glyphosate) K_M values

SEQ ID NO.	Clone ID	K_M(mM)
SEQ ID NO:263	13_10F6	1.3
SEQ ID NO:264	13_12G6	1.2
SEQ ID NO:265	14_2A5	1.6
SEQ ID NO:266	14_2C1	3.1
SEQ ID NO:267	14_2F11	1.7
SEQ ID NO:268	CHIMERA	1.3
SEQ ID NO:269	10_12D7	1.8
SEQ ID NO:270	10_15F4	1
SEQ ID NO:271	10_17D1	2.2
SEQ ID NO:272	10_17F6	1.4
SEQ ID NO:273	10_18G9	1.2
SEQ ID NO:274	10_1H3	1.9
SEQ ID NO:275	10_20D10	1.6
SEQ ID NO:276	10_23F2	0.9
SEQ ID NO:277	10_2B8	1.1
SEQ ID NO:278	10_2C7	1.4
SEQ ID NO:279	10_3G5	2
SEQ ID NO:280	10_4H7	1.7
SEQ ID NO:281	10_6D11	1.2
SEQ ID NO:282	10_8C6	0.7
SEQ ID NO:283	11C3	3.1

SEQ ID NO:284	11G3	1.7
SEQ ID NO:285	11H3	1.4
SEQ ID NO:286	12_1F9	3
SEQ ID NO:287	12_2G9	1.5
SEQ ID NO:288	12_3F1	0.9
SEQ ID NO:289	12_5C10	1.5
SEQ ID NO:290	12_6A10	1.1
SEQ ID NO:291	12_6D1	1.2
SEQ ID NO:292	12_6F9	1.9
SEQ ID NO:293	12_6H6	1.6
SEQ ID NO:294	12_7D6	1.4
SEQ ID NO:295	12_7G11	2
SEQ ID NO:296	12F5	1.8
SEQ ID NO:297	12G7	3.7
SEQ ID NO:298	1_2H6	0.9
SEQ ID NO:299	13_12G12	0.69
SEQ ID NO:300	13_6D10	0.65
SEQ ID NO:301	13_7A7	0.5
SEQ ID NO:302	13_7B12	1.7
SEQ ID NO:303	13_7C1	1.5
SEQ ID NO:304	13_8G6	0.61
SEQ ID NO:305	13_9F6	1.3
SEQ ID NO:306	14_10C9	0.9
SEQ ID NO:307	14_10H3	0.6
SEQ ID NO:308	14_10H9	1.1
SEQ ID NO:309	14_11C2	1
SEQ ID NO:310	14_12D8	1
SEQ ID NO:311	14_12H6	0.9
SEQ ID NO:312	14_2B6	0.63
SEQ ID NO:313	14_2G11	1.4
SEQ ID NO:314	14_3B2	0.85
SEQ ID NO:315	14_4H8	2
SEQ ID NO:316	14_6A8	0.78
SEQ ID NO:317	14_6B10	1.4
SEQ ID NO:318	14_6D4	1
SEQ ID NO:319	14_7A11	3.7
SEQ ID NO:320	14_7A1	1.6
SEQ ID NO:321	14_7A9	3.2
SEQ ID NO:322	14_7G1	0.66
SEQ ID NO:323	14_7H9	1.3
SEQ ID NO:324	14_8F7	1.8
SEQ ID NO:325	15_10C2	0.8
SEQ ID NO:326	15_10D6	1
SEQ ID NO:327	15_11F9	1
SEQ ID NO:328	15_11H3	1
SEQ ID NO:329	15_12A8	1.6
SEQ ID NO:330	15_12D6	0.74
SEQ ID NO:331	15_12D8	1.3
SEQ ID NO:332	15_12D9	1.4
SEQ ID NO:333	15_3F10	0.9
SEQ ID NO:334	15_3G11	1.2
SEQ ID NO:335	15_4F11	0.9
SEQ ID NO:336	15_4H3	1
SEQ ID NO:337	15_6D3	1.4
SEQ ID NO:338	15_6G11	0.9
SEQ ID NO:339	15_9F6	1.1
SEQ ID NO:340	15F5	2.9
SEQ ID NO:341	16A1	2.9

SEQ ID NO:342	16H3	2.9
SEQ ID NO:343	17C12	1.4
SEQ ID NO:344	18D6	1.2
SEQ ID NO:345	19C6	1.1
SEQ ID NO:346	19D5	1.7
SEQ ID NO:347	20A12	1.1
SEQ ID NO:348	20F2	1.9
SEQ ID NO:349	2.10E+12	0.7
SEQ ID NO:350	23H11	2.2
SEQ ID NO:351	24C1	0.9
SEQ ID NO:352	24C6	1.3
SEQ ID NO:353	2.40E+08	0.9
SEQ ID NO:354	2_8C3	1.5
SEQ ID NO:355	2H3	0.9
SEQ ID NO:356	30G8	1.6
SEQ ID NO:357	3B_10C4	1.6
SEQ ID NO:358	3B_10G7	1
SEQ ID NO:359	3B_12B1	1.2
SEQ ID NO:360	3B_12D10	0.9
SEQ ID NO:361	3B_2E5	1.3
SEQ ID NO:362	3C_10H3	1.1
SEQ ID NO:363	3C_12H10	1.2
SEQ ID NO:364	3C_9H8	1
SEQ ID NO:365	4A_1B11	1.6
SEQ ID NO:366	4A_1C2	1.2
SEQ ID NO:367	4B_13E1	2
SEQ ID NO:368	4B_13G10	7.6
SEQ ID NO:369	4B_16E1	1
SEQ ID NO:370	4B_17A1	1.1
SEQ ID NO:371	4B_18F11	1.7
SEQ ID NO:372	4B_19C8	1.2
SEQ ID NO:373	4B_1G4	1
SEQ ID NO:374	4B_21C6	0.8
SEQ ID NO:375	4B_2H7	6.2
SEQ ID NO:376	4B_2H8	1.2
SEQ ID NO:377	4B_6D8	1.5
SEQ ID NO:378	4B_7E8	1.2
SEQ ID NO:379	4C_8C9	0.6
SEQ ID NO:380	4H1	1.4
SEQ ID NO:381	6_14D10	1.5
SEQ ID NO:382	6_15G7	1.3
SEQ ID NO:383	6_16A5	1.1
SEQ ID NO:384	6_16F5	1
SEQ ID NO:385	6_17C5	1.3
SEQ ID NO:386	6_18C7	1.2
SEQ ID NO:387	6_18D7	1.2
SEQ ID NO:388	6_19A10	1.9
SEQ ID NO:389	6_19B6	0.7
SEQ ID NO:390	6_19C3	1.4
SEQ ID NO:391	6_19C8	2
SEQ ID NO:392	6_20A7	1
SEQ ID NO:393	6_20A9	1.3
SEQ ID NO:394	6_20H5	0.8
SEQ ID NO:395	6_21F4	0.7
SEQ ID NO:396	6_22C9	3.2
SEQ ID NO:397	6_22D9	1.3
SEQ ID NO:398	6_22H9	1.1
SEQ ID NO:399	6_23H3	1.1

SEQ ID NO:400	6_23H7	1.2
SEQ ID NO:401	6_2H1	0.9
SEQ ID NO:402	6_3D6	1
SEQ ID NO:403	6_3G3	1
SEQ ID NO:404	6_3H2	1
SEQ ID NO:405	6_4A10	1.1
SEQ ID NO:406	6_4B1	1
SEQ ID NO:407	6_5D11	1
SEQ ID NO:408	6_5F11	1.9
SEQ ID NO:409	6_5G9	1.4
SEQ ID NO:410	6_6D5	1
SEQ ID NO:411	6_7D1	0.5
SEQ ID NO:412	6_8H3	1
SEQ ID NO:413	6_9G11	1.3
SEQ ID NO:414	6F1	1.8
SEQ ID NO:415	7_1C4	1.1
SEQ ID NO:416	7_2A10	0.8
SEQ ID NO:417	7_2A11	1.1
SEQ ID NO:418	7_2D7	1.1
SEQ ID NO:419	7_5C7	1
SEQ ID NO:420	7_9C9	1
SEQ ID NO:421	9_13F10	0.7
SEQ ID NO:422	9_13F1	1.1
SEQ ID NO:423	9_15D5	1.2
SEQ ID NO:424	9_15D8	1.1
SEQ ID NO:425	9_15H3	1.9
SEQ ID NO:426	9_18H2	1.1
SEQ ID NO:427	9_20F12	1
SEQ ID NO:428	9_21C8	1.2
SEQ ID NO:429	9_22B1	1.4
SEQ ID NO:430	9_23A10	1
SEQ ID NO:431	9_24F6	0.9
SEQ ID NO:432	9_4H10	1.5
SEQ ID NO:433	9_4H8	0.6
SEQ ID NO:434	9_8H1	1.7
SEQ ID NO:435	9_9H7	0.7
SEQ ID NO:436	9C6	2.5
SEQ ID NO:437	9H11	2.3
SEQ ID NO:438	0_4B10	0.68
SEQ ID NO:439	0_5B11	0.54
SEQ ID NO:440	0_5B3	0.39
SEQ ID NO:441	0_5B4	0.6
SEQ ID NO:442	0_5B8	0.27
SEQ ID NO:443	0_5C4	0.67
SEQ ID NO:444	0_5D11	0.67
SEQ ID NO:445	0_5D3	0.5
SEQ ID NO:446	0_5D7	1.1
SEQ ID NO:447	0_6B4	0.8
SEQ ID NO:448	0_6D10	0.1
SEQ ID NO:449	0_6D11	0.44
SEQ ID NO:450	0_6F2	0.34
SEQ ID NO:451	0_6H9	0.47
SEQ ID NO:452	10_4C10	0.1
SEQ ID NO:453	10_4D5	0.1
SEQ ID NO:454	10_4F2	0.2
SEQ ID NO:455	10_4F9	0.1
SEQ ID NO:456	10_4G5	0.58
SEQ ID NO:457	10_4H4	0.1

SEQ ID NO:458	11_3A11	0.1
SEQ ID NO:459	11_3B1	0.63
SEQ ID NO:460	11_3B5	0.26
SEQ ID NO:461	11_3C12	0.1
SEQ ID NO:462	11_3C3	0.22
SEQ ID NO:463	11_3C6	0.21
SEQ ID NO:464	11_3D6	0.1
SEQ ID NO:465	1_1G12	0.1
SEQ ID NO:466	1_1H1	1.8
SEQ ID NO:467	1_1H2	0.44
SEQ ID NO:468	1_1H5	1.5
SEQ ID NO:469	1_2A12	1.3
SEQ ID NO:470	1_2B6	0.58
SEQ ID NO:471	1_2C4	0.8
SEQ ID NO:472	1_2D2	1.2
SEQ ID NO:473	1_2D4	1.2
SEQ ID NO:474	1_2F8	1.9
SEQ ID NO:475	1_2H8	0.48
SEQ ID NO:476	1_3A2	0.8
SEQ ID NO:477	1_3D6	3.5
SEQ ID NO:478	1_3F3	1.5
SEQ ID NO:479	1_3H2	0.7
SEQ ID NO:480	1_4C5	0.93
SEQ ID NO:481	1_4D6	1.4
SEQ ID NO:482	1_4H1	1.2
SEQ ID NO:483	1_5H5	0.51
SEQ ID NO:484	1_6F12	14.7
SEQ ID NO:485	1_6H6	1.05
SEQ ID NO:486	3_11A10	0.17
SEQ ID NO:487	3_14F6	0.25
SEQ ID NO:488	3_15B2	0.1
SEQ ID NO:489	3_6A10	0.66
SEQ ID NO:490	3_6B1	0.43
SEQ ID NO:491	3_7F9	0.29
SEQ ID NO:492	3_8G11	0.1
SEQ ID NO:493	4_1B10	0.1
SEQ ID NO:494	5_2B3	0.1
SEQ ID NO:495	5_2D9	0.1
SEQ ID NO:496	5_2F10	0.1
SEQ ID NO:497	6_1A11	0.1
SEQ ID NO:498	6_1D5	0.1
SEQ ID NO:499	6_1F11	0.1
SEQ ID NO:500	6_1F1	0.1
SEQ ID NO:501	6_1H10	0.1
SEQ ID NO:502	6_1H4	0.1
SEQ ID NO:503	8_1F8	0.1
SEQ ID NO:504	8_1G2	0.1
SEQ ID NO:505	8_1G3	0.1
SEQ ID NO:506	8_1H7	0.1
SEQ ID NO:507	8_1H9	0.1
SEQ ID NO:508	GAT1_21F12	4.6
SEQ ID NO:509	GAT1_24G3	3.8
SEQ ID NO:510	GAT1_29G1	4
SEQ ID NO:511	GAT1_32G1	3.3
SEQ ID NO:512	GAT2_15G8	2.8
SEQ ID NO:513	GAT2_19H8	2.8
SEQ ID NO:514	GAT2_21F1	3

Table 5. GAT polypeptide k_{cat}/ K_M values

SEQ ID NO.	Clone ID	$K_{cat}/K_M(mM^{-1} min^{-1})$
SEQ ID NO:263	13_10F6	37.4
SEQ ID NO:264	13_12G6	43.4
SEQ ID NO:265	14_2A5	175.5
SEQ ID NO:266	14_2C1	43
SEQ ID NO:267	14_2F11	80.6
SEQ ID NO:268	CHIMERA	119.6
SEQ ID NO:269	10_12D7	43
SEQ ID NO:270	10_15F4	37.6
SEQ ID NO:271	10_17D1	80.1
SEQ ID NO:272	10_17F6	34.2
SEQ ID NO:273	10_18G9	20
SEQ ID NO:274	10_1H3	40.1
SEQ ID NO:275	10_20D10	53.9
SEQ ID NO:276	10_23F2	112.5
SEQ ID NO:277	10_2B8	98.5
SEQ ID NO:278	10_2C7	96.4
SEQ ID NO:279	10_3G5	43.7
SEQ ID NO:280	10_4H7	65.9
SEQ ID NO:281	10_6D11	52
SEQ ID NO:282	10_8C6	31
SEQ ID NO:283	11C3	0.9
SEQ ID NO:284	11G3	8.9
SEQ ID NO:285	11H3	0.9
SEQ ID NO:286	12_1F9	26.8
SEQ ID NO:287	12_2G9	101
SEQ ID NO:288	12_3F1	49
SEQ ID NO:289	12_5C10	59.7
SEQ ID NO:290	12_6A10	49.7
SEQ ID NO:291	12_6D1	40.8
SEQ ID NO:292	12_6F9	46.9
SEQ ID NO:293	12_6H6	56.5
SEQ ID NO:294	12_7D6	38.5
SEQ ID NO:295	12_7G11	117.2
SEQ ID NO:296	12F5	1.7
SEQ ID NO:297	12G7	0.6
SEQ ID NO:298	1_2H6	10.4
SEQ ID NO:299	13_12G12	52.4
SEQ ID NO:300	13_6D10	456.1
SEQ ID NO:301	13_7A7	234
SEQ ID NO:302	13_7B12	40.5
SEQ ID NO:303	13_7C1	32.1
SEQ ID NO:304	13_8G6	55.2
SEQ ID NO:305	13_9F6	45.3
SEQ ID NO:306	14_10C9	141.1
SEQ ID NO:307	14_10H3	175.3
SEQ ID NO:308	14_10H9	115.6
SEQ ID NO:309	14_11C2	108.7
SEQ ID NO:310	14_12D8	62.1
SEQ ID NO:311	14_12H6	101.3
SEQ ID NO:312	14_2B6	54.3
SEQ ID NO:313	14_2G11	49.6
SEQ ID NO:314	14_3B2	80.9

SEQ ID NO:315	14_4H8	99.4
SEQ ID NO:316	14_6A8	56
SEQ ID NO:317	14_6B10	96.2
SEQ ID NO:318	14_6D4	256
SEQ ID NO:319	14_7A11	53.3
SEQ ID NO:320	14_7A1	97.4
SEQ ID NO:321	14_7A9	76.9
SEQ ID NO:322	14_7G1	207.1
SEQ ID NO:323	14_7H9	49.5
SEQ ID NO:324	14_8F7	50.3
SEQ ID NO:325	15_10C2	87.3
SEQ ID NO:326	15_10D6	67.1
SEQ ID NO:327	15_11F9	76.4
SEQ ID NO:328	15_11H3	61.9
SEQ ID NO:329	15_12A8	48.2
SEQ ID NO:330	15_12D6	200.8
SEQ ID NO:331	15_12D8	45.9
SEQ ID NO:332	15_12D9	42.6
SEQ ID NO:333	15_3F10	54.1
SEQ ID NO:334	15_3G11	59.6
SEQ ID NO:335	15_4F11	89.2
SEQ ID NO:336	15_4H3	93.3
SEQ ID NO:337	15_6D3	61.3
SEQ ID NO:338	15_6G11	41
SEQ ID NO:339	15_9F6	54.2
SEQ ID NO:340	15F5	0.2
SEQ ID NO:341	16A1	3.6
SEQ ID NO:342	16H3	1.2
SEQ ID NO:343	17C12	2.3
SEQ ID NO:344	18D6	8
SEQ ID NO:345	19C6	2
SEQ ID NO:346	19D5	1.3
SEQ ID NO:347	20A12	2.5
SEQ ID NO:348	20F2	2
SEQ ID NO:349	2.10E+12	1.5
SEQ ID NO:350	23H11	3.2
SEQ ID NO:351	24C1	1.8
SEQ ID NO:352	24C6	2.1
SEQ ID NO:353	2.40E+08	9.8
SEQ ID NO:354	2_8C3	16.6
SEQ ID NO:355	2H3	17.7
SEQ ID NO:356	30G8	6.4
SEQ ID NO:357	3B_10C4	15.5
SEQ ID NO:358	3B_10G7	19.6
SEQ ID NO:359	3B_12B1	19
SEQ ID NO:360	3B_12D10	6
SEQ ID NO:361	3B_2E5	12.6
SEQ ID NO:362	3C_10H3	30.8
SEQ ID NO:363	3C_12H10	7.6
SEQ ID NO:364	3C_9H8	11.7
SEQ ID NO:365	4A_1B11	15
SEQ ID NO:366	4A_1C2	17
SEQ ID NO:367	4B_13E1	18.6
SEQ ID NO:368	4B_13G10	4.6
SEQ ID NO:369	4B_16E1	17
SEQ ID NO:370	4B_17A1	17.4
SEQ ID NO:371	4B_18F11	8.6
SEQ ID NO:372	4B_19C8	13.2

SEQ ID NO:373	4B_1G4	3.7
SEQ ID NO:374	4B_21C6	14.8
SEQ ID NO:375	4B_2H7	4.4
SEQ ID NO:376	4B_2H8	31.9
SEQ ID NO:377	4B_6D8	15.2
SEQ ID NO:378	4B_7E8	17.1
SEQ ID NO:379	4C_8C9	15.1
SEQ ID NO:380	4H1	0.9
SEQ ID NO:381	6_14D10	28.2
SEQ ID NO:382	6_15G7	37.3
SEQ ID NO:383	6_16A5	39.8
SEQ ID NO:384	6_16F5	35.2
SEQ ID NO:385	6_17C5	27.1
SEQ ID NO:386	6_18C7	26.8
SEQ ID NO:387	6_18D7	35.8
SEQ ID NO:388	6_19A10	45.7
SEQ ID NO:389	6_19B6	34.2
SEQ ID NO:390	6_19C3	16.5
SEQ ID NO:391	6_19C8	37.4
SEQ ID NO:392	6_20A7	40.4
SEQ ID NO:393	6_20A9	34.7
SEQ ID NO:394	6_20H5	24.3
SEQ ID NO:395	6_21F4	34.7
SEQ ID NO:396	6_22C9	14.8
SEQ ID NO:397	6_22D9	33.8
SEQ ID NO:398	6_22H9	15.9
SEQ ID NO:399	6_23H3	39.9
SEQ ID NO:400	6_23H7	38.5
SEQ ID NO:401	6_2H1	29.5
SEQ ID NO:402	6_3D6	41.7
SEQ ID NO:403	6_3G3	51.9
SEQ ID NO:404	6_3H2	57.2
SEQ ID NO:405	6_4A10	50
SEQ ID NO:406	6_4B1	27
SEQ ID NO:407	6_5D11	15.2
SEQ ID NO:408	6_5F11	21.1
SEQ ID NO:409	6_5G9	25.6
SEQ ID NO:410	6_6D5	55.3
SEQ ID NO:411	6_7D1	39.5
SEQ ID NO:412	6_8H3	44.7
SEQ ID NO:413	6_9G11	60.3
SEQ ID NO:414	6F1	5.6
SEQ ID NO:415	7_1C4	15.9
SEQ ID NO:416	7_2A10	18.2
SEQ ID NO:417	7_2A11	42.6
SEQ ID NO:418	7_2D7	49.9
SEQ ID NO:419	7_5C7	44.7
SEQ ID NO:420	7_9C9	65
SEQ ID NO:421	9_13F10	49.6
SEQ ID NO:422	9_13F1	28.7
SEQ ID NO:423	9_15D5	23
SEQ ID NO:424	9_15D8	97.6
SEQ ID NO:425	9_15H3	36.2
SEQ ID NO:426	9_18H2	22.7
SEQ ID NO:427	9_20F12	37.8
SEQ ID NO:428	9_21C8	23.8
SEQ ID NO:429	9_22B1	35.8
SEQ ID NO:430	9_23A10	21

SEQ ID NO:431	9_24F6	58.3
SEQ ID NO:432	9_4H10	67.5
SEQ ID NO:433	9_4H8	78.5
SEQ ID NO:434	9_8H1	44
SEQ ID NO:435	9_9H7	40
SEQ ID NO:436	9C6	5.1
SEQ ID NO:437	9H11	1.7
SEQ ID NO:438	0_4B10	279
SEQ ID NO:439	0_5B11	406
SEQ ID NO:440	0_5B3	367
SEQ ID NO:441	0_5B4	301
SEQ ID NO:442	0_5B8	522
SEQ ID NO:443	0_5C4	306
SEQ ID NO:444	0_5D11	334
SEQ ID NO:445	0_5D3	660
SEQ ID NO:446	0_5D7	222
SEQ ID NO:447	0_6B4	315
SEQ ID NO:448	0_6D10	1177
SEQ ID NO:449	0_6D11	481
SEQ ID NO:450	0_6F2	516
SEQ ID NO:451	0_6H9	486
SEQ ID NO:452	10_4C10	695.98
SEQ ID NO:453	10_4D5	827.16
SEQ ID NO:454	10_4F2	1155.19
SEQ ID NO:455	10_4F9	553.93
SEQ ID NO:456	10_4G5	304.57
SEQ ID NO:457	10_4H4	1183.6
SEQ ID NO:458	11_3A11	556.62
SEQ ID NO:459	11_3B1	349.17
SEQ ID NO:460	11_3B5	748.49
SEQ ID NO:461	11_3C12	490.67
SEQ ID NO:462	11_3C3	972.81
SEQ ID NO:463	11_3C6	878.27
SEQ ID NO:464	11_3D6	553.01
SEQ ID NO:465	1_1G12	584.79
SEQ ID NO:466	1_1H1	162
SEQ ID NO:467	1_1H2	366
SEQ ID NO:468	1_1H5	63
SEQ ID NO:469	1_2A12	176
SEQ ID NO:470	1_2B6	239
SEQ ID NO:471	1_2C4	242
SEQ ID NO:472	1_2D2	104
SEQ ID NO:473	1_2D4	152
SEQ ID NO:474	1_2F8	85
SEQ ID NO:475	1_2H8	294
SEQ ID NO:476	1_3A2	227
SEQ ID NO:477	1_3D6	64
SEQ ID NO:478	1_3F3	112
SEQ ID NO:479	1_3H2	183
SEQ ID NO:480	1_4C5	273
SEQ ID NO:481	1_4D6	98
SEQ ID NO:482	1_4H1	196
SEQ ID NO:483	1_5H5	419
SEQ ID NO:484	1_6F12	14
SEQ ID NO:485	1_6H6	259
SEQ ID NO:486	3_11A10	796.55
SEQ ID NO:487	3_14F6	753.73
SEQ ID NO:488	3_15B2	1041.32

SEQ ID NO:489	3_6A10	191.64
SEQ ID NO:490	3_6B1	611.81
SEQ ID NO:491	3_7F9	667.4
SEQ ID NO:492	3_8G11	991.44
SEQ ID NO:493	4_1B10	770.91
SEQ ID NO:494	5_2B3	567.5
SEQ ID NO:495	5_2D9	754.36
SEQ ID NO:496	5_2F10	547.22
SEQ ID NO:497	6_1A11	455.41
SEQ ID NO:498	6_1D5	429.16
SEQ ID NO:499	6_1F11	1057.6
SEQ ID NO:500	6_1F1	698.15
SEQ ID NO:501	6_1H10	170.11
SEQ ID NO:502	6_1H4	859.12
SEQ ID NO:503	8_1F8	828.78
SEQ ID NO:504	8_1G2	674.73
SEQ ID NO:505	8_1G3	1088.97
SEQ ID NO:506	8_1H7	1012.4
SEQ ID NO:507	8_1H9	783.89
SEQ ID NO:508	GAT1_21F12	1.2
SEQ ID NO:509	GAT1_24G3	1.3
SEQ ID NO:510	GAT1_29G1	1.5
SEQ ID NO:511	GAT1_32G1	1.4
SEQ ID NO:512	GAT2_15G8	1.6
SEQ ID NO:513	GAT2_19H8	1.5
SEQ ID NO:514	GAT2_21F1	1.4

K_M for AcCoA is measured using the mass spectrometry method with repeated sampling during the reaction. Acetyl-coenzyme A and glyphosate (ammonium salts) are placed as 50-fold-concentrated stock solutions into a well of a mass spectrometry sample plate. Reactions are initiated with the addition of enzyme appropriately diluted in a volatile buffer such as morpholine acetate or ammonium carbonate, pH 6.8 or 7.7. The sample is repeatedly injected into the instrument and initial rates are calculated from plots of retention time and peak area. K_M is calculated as for glyphosate.

EXAMPLE 8: SELECTION OF TRANSFORMED E. COLI

An evolved *gat* gene (a chimera with a native *B. licheniformis* ribosome binding site (AACTGAAGGAGGAATCTC; SEQ ID NO:515) attached directly to the 5' end of the GAT coding sequence) was cloned into the expression vector pQE80 (Qiagen) between the EcoRI and HindIII sites, resulting in the plasmid pMAXY2190 (Figure 11). This eliminated the His tag domain from the plasmid and retained the B-lactamase gene conferring resistance to the antibiotics ampicillin and carbenicillin. pMAXY2190 was electroporated (BioRad Gene Pulser) into XL1 Blue (Stratagene) *E. coli* cells. The cells were suspended in SOC rich medium and allowed to recover for one hour. The cells were

then gently pelleted, washed one time with M9 minimal media lacking aromatic amino acids (12.8 g/L Na2HPO4.7 H2O, 3.0 g/L KH2PO4, 0.5 g/L NaCl, 1.0 g/L NH4Cl, 0.4% glucose, 2 mM MgSO4, 0.1 mM CaCl2, 10 mg/L thiamine, 10 mg/L proline, 30 mg/L carbenicillin), and resuspended in 20 ml of the same M9 medium. After overnight growth at 37°C at 250 rpm, equal volumes of cells were plated on either M9 medium or M9 plus 1 mM glyphosate medium. pQE80 vector with no *gat* gene was similarly introduced into *E. coli* cells and plated for single colonies for comparison. The results are summarized in Table 6 and clearly demonstrate that GAT activity allows selection and growth of transformed *E. coli* cells with less than 1% background. Note that no IPTG induction was necessary for sufficient GAT activity to allow growth of transformed cells. Transformation was verified by re-isolation of pMAXY2190 from the *E. coli* cells grown in the presence of glyphosate.

Table 6. Glyphosate selection of pMAXY2190 in *E. coli*

	Number of colonies	
Plasmid	M9 - glyphosate	M9 + 1 mM glyphosate
pMAXY2190	568	512
pQE80	324	3

EXAMPLE 9: SELECTION OF TRANSFORMED PLANT CELLS

Agrobacterium-mediated transformation of plant cells occurs at low efficiencies. To allow propagation of transformed cells while inhibiting proliferation of non-transformed cells, a selectable marker is needed. Antibiotic markers for kanamycin and hygromycin and the herbicide modifying gene *bar*, which detoxifies the herbicidal compound phosphinothricin, are examples of selectable markers used in plants (Methods in Molecular Biology, 1995, 49:9-18). Here we demonstrate that GAT activity serves as an efficient selectable marker for plant transformation. An evolved *gat* gene (0_5B8) was cloned between a plant promoter (enhanced strawberry vein banded virus) and a ubiquinone terminator and introduced into the T-DNA region of the binary vector pMAXY3793 suitable for transformation of plant cells via *Agrobacterium tumefaciens* EHA105 as shown in Figure 12. A screenable GUS marker was present in the T-DNA to allow confirmation of transformation. Transgenic tobacco shoots were generated using glyphosate as the only selecting agent.

Axillary buds of *Nicotiana tabacum* L. Xanthi were subcultured on half-strength MS medium with sucrose (1.5 %) and Gelrite (0.3 %) under 16-h light (35-42

μEinsteins m^{-2} s^{-1}, cool white fluorescent lamps) at 24 °C every 2-3 weeks. Young leaves were excised from plants after 2-3 weeks subculture and were cut into 3 x 3 mm segments. *A. tumefaciens* EHA105 was inoculated into LB medium and grown overnight to a density of A600= 1.0. Cells were pelleted at 4,000 rpm for 5 minutes and resuspended in 3 volumes of liquid co-cultivation medium composed of Murashige and Skoog (MS) medium (pH 5.2) with 2 mg/L N6-benzyladenine (BA), 1% glucose and 400 uM acetysyringone. The leaf pieces were then fully submerged in 20 ml of *A. tumefaciens* in 100 x 25 mm Petri dishes for 30 min, blotted with autoclaved filter paper, then placed on solid co-cultivation medium (0.3% Gelrite) and incubated as described above. After 3 days of co-cultivation, 20-30 segments were transferred to basal shoot induction (BSI) medium composed of MS solid medium (pH 5.7) with 2 mg/L BA, 3% sucrose, 0.3% Gelrite, 0-200 uM glyphosate, and 400 ug/ml Timentin.

After 3 weeks, shoots were clearly evident on the explants placed on media with no glyphosate regardless of the presence or absence of the *gat* gene. T-DNA transfer from both constructs was confirmed by GUS histochemical staining of leaves from regenerated shoots. Glyphosate concentrations greater than 20 uM completely inhibited any shoot formation from the explants lacking a *gat* gene. Explants infected with *A. tumefaciens* with the gat construct regenerated shoots at glyphosate concentrations up to 200 uM (the highest level tested). Transformation was confirmed by GUS histochemical staining and by PCR fragment amplification of the *gat* gene using primers annealing to the promoter and 3' regions. The results are summarized in Table 7.

Table 7. Tobacco shoot regeneration with glyphosate selection.

Glyphosate concentration % Shoot Regeneration					
Transferred genes	0 uM	20 uM	40 uM	80 uM	200 uM
GUS	100	0	0	0	0
gat and GUS	100	60	30	5	3

EXAMPLE 10: GLYPHOSATE SELECTION OF TRANSFORMED YEAST CELLS

Selection markers for yeast transformation are usually auxotrophic genes that allow growth of transformed cells on a medium lacking the specific amino acid or nucleotide. Because *Saccharomyces cerevisiae* is sensitive to glyphosate, GAT can also be used as a selectable marker. To demonstrate this, an evolved *gat* gene (0_6D10) is cloned from the T-DNA vector pMAXY3793 (as shown in Example 9) as a PstI-ClaI fragment containing the entire coding region and ligated into PstI-ClaI digested p424TEF (Gene, 1995, 156:119-122) as shown in Figure 13. This plasmid contains an *E. coli* origin of replication and a gene conferring carbenicillin resistance as well as a TRP1, tryptophan auxotroph selectable marker for yeast transformation.

The *gat* containing construct is transformed into *E. coli* XL1 Blue (Statagene) and plated on LB carbenicillin (50 ug/ml) agar medium. Plasmid DNA is prepared and used to transform yeast strain YPH499 (Stratagene) using a transformation kit (Bio101). Equal amounts of transformed cells are plated on CSM-YNB-glucose medium (Bio101) lacking all aromatic amino acids (tryptophan, tyrosine, and phenylalanine) with added glyphosate. For comparison, p424TEF lacking the *gat* gene is also introduced into YPH499 and plated as described. The results demonstrate that GAT activity function will as an efficient selectable marker. The presence of the *gat* containing vector in glyphosate selected colonies can be confirmed by re-isolation of the plasmid and restriction digest analysis.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques, methods, compositions, apparatus and systems described above may be used in various combinations. The invention is intended to include all methods and reagents described herein, as well as all polynculeotides, polypeptides, cells, organisms, plants, crops, etc., that are the products of these novel methods and reagents.

All publications, patents, patent applications, or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, or other document were individually indicated to be incorporated by reference for all purposes.

SEQ ID NO.	Clone ID	Sequence
SEQ ID NO:1	ST401 gat	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCTT GAAGGCGAAGAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGATACCGTGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAATGCCAGGACATCTGTG AGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:2	B6 gat	ATGATTGAAGTCAAACCTATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTATCGGGACAGGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGATATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACATAA
SEQ ID NO:3	DS3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATAATGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGAAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGATATCTGTG AGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAAGG CGGGATCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGGCATAA
SEQ ID NO:4	NHA-2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATAATGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGAAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGATATCTGTG AGCGGCTACTATGAAAAGCTCGGCCTCAGCGAACAAGG CGGGATCTACGACATACCGCCGATCGGACCTCATATTTT

		GATGTATAAGAAATTGGCATAA
SEQ ID NO:5	NH5-2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCTT GAGGGCGAAGAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGATACCGTGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAATGCCAGGACATCTGTG AGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:6	ST401 GAT	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGAFH LGGYYRGKLISIASFHKAEHSELEGEEQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:7	B6 GAT	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRDRLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:8	DS3 GAT	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHNAEHSELEGQKQYQLRGMA TLEGYREQKAGSTLIRHAEELLRKKGADLLWCNARISVSG YYEKLGFSEQGGIYDIPPIGPHILMYKKLA
SEQ ID NO:9	NHA-2 GAT	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHNAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARISVSGYYEKL GLSEQGGIYDIPPIGPHILMYKKLA
SEQ ID NO:10	NH5-2 GAT	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGAFH LGGYYQGKLISIASFHKAEHSELEGEEQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:11	13_10F6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACGTCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:12	13_12G6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT

		GAAGGCCAAAGACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACTGGGCCCCATATTTTG ATGTATAAGAAATTGACATAA
SEQ ID NO:13	14_2A5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGAGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGAAGCAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACGTCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACACACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:14	14_2C1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGACTGGGCCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:15	14_2F11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTTGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGCCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:16	CHIMERA	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGCGAGCAAAAAGCGGGCAGTACG

		CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:17	10_12D7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGNATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGCTTCACCTCGGTGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAG GCGAAGTCTACGACATACCGCCGACCGGACCCCATATT TTGATGTATAAGAAATTGACGTAA
SEQ ID NO:18	10_15F4	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCAGGGGT ACGTTTCACCTCGGTGGGTATTACCGGGGCAAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:19	10_17D1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG CGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:20	10_17F6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GCGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG

		AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:21	10_18G9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACTGATTTGCTCGGTGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:22	10_1H3	ATGATTGAAGTCAAACCAATAAACGCGGAGGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTATCGGGGCAAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCGAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACATAA
SEQ ID NO:23	10_20D10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGCTTCACCTCGGTGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:24	10_23F2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACACCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG

		ATGTATAAGAAATTGACGTAA
SEQ ID NO:25	10_2B8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:26	10_2C7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTCGAAGGGTACCGTGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAA GGCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACACACCGCCGGTCGGACCTCATATT TTGATGTATAAGAAATTGACGTAA
SEQ ID NO:27	10_3G5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:28	10_4H7	ATGATTGAAGTCAAACCGATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID	10_6D11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA

NO:29		TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGCTTCACCTCGGTGGATATTACCGGGGCAAGCTGGT CAGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:30	10_8C6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:31	11C3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGATATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACATAA
SEQ ID NO:32	11G3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCGTGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCAGGGCAAGCTGAT CAGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGATATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGGCATAA
SEQ ID NO:33	11H3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT

		GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACACCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCAACTGGGCCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:34	12_1F9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCCACGACATACCGCCGACCGGACCCCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:35	12_2G9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACACACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:36	12_3F1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTTGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGAAGTAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:37	12_5C10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTATCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT

		GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACGCACCGCCGACCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:38	12_6A10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:39	12_6D1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCTGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:40	12_6F9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:41	12_6H6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCACCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG

		CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACATAA
SEQ ID NO:42	12_7D6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACTGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:43	12_7G11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:44	12F5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCTT GAGGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAATGCCAGGACATCTGTGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGATCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:45	12G7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGATACCGTGAGCAAAAAGCGGGAAGCAC ACTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTG

		AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:46	1_2H6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:47	13_12G12	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGCATAAGAAATTGACGTAA
SEQ ID NO:48	13_6D10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTCGCTCGGAGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGGTACCGTGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACACACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:49	13_7A7	ATGATCGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCAGGAGT GCGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCACCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGGGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGTA CGCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACACACCGCCGGTCGGACCTCATATT

		TTGATGTATAAGAAATTGACGTAA
SEQ ID NO:50	13_7B12	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGAGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACTGGGCCCCATATTTT GATGTATAAGAAGTTGACGTAA
SEQ ID NO:51	13_7C1	ATGATTGAAGTCAAACCAATAAATGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TTGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAACTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGTAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCGA GAGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GAAGTCTACGACATACCGCCGACTGGGCCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:52	13_8G6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTCGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCT TGAAGGTCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACGTCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:53	13_9F6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATCTGCTTGGGGGC ACGTTTCACCTAGGTGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGTA CGCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGGTCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID	14_10C9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA

NO:54		TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TAGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCTGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGTAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACGTCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACACACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAGTTGACGTAA
SEQ ID NO:55	14_10H3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAA GGCGCAGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACACACCGCCGGTCGGACCTCATATT TTGATGTATAAGAAGTTGACGTAA
SEQ ID NO:56	14_10H9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACATAA
SEQ ID NO:57	14_11C2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGAGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACACACCGCCGACCGGACCCCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:58	14_12D8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTAAGTATGAAACCGATTTGCTCGGGGGT

		ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGATACCGTGAGCAAAAAGCTGGCAGTAC GCTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAAG GCGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCG AGCGGCTACTATAAAAAGCTCGGCTTCAGGGAACAAGG CGGGGTCTACGACATACCGCCTGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:59	14_12H6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTGTGGTGCAACGCCAGGACGTCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACTGGGCCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:60	14_2B6	ATGATTGAAGTCAAACCAATAAATGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACGTCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:61	14_2G11	ATGATTGAAGTCAAACCAATAAATGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCGA GTGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACTGGGCCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:62	14_3B2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCAGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAGGCCGAACATCCAGAGCTT

		GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGAAGCAC GCTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGCCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:63	14_4H8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGAGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTCGAAGGGTACCGTGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAA GGCGCGGACCTTTTGTGGTGCAACGCCAGGACGTCTGC GAGCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACACACCGCCGGTCGGACCTCATATT TTGATGTATAAGAAATTGACGTAA
SEQ ID NO:64	14_6A8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTAGTC AGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATGTTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:65	14_6B10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTTGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATGCCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAGTTGACGTAA
SEQ ID NO:66	14_6D4	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGACCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGAGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG

		CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:67	14_7A11	ATGATTGAAGTCAAACCAATAAACGCGGAGGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCTAAAACAGTATCAGCTGAGAGGGATGGCGAC ACTCGAAGGGTACCGTGAGCAAAAAGCGGGAAGTACG CTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACGTCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGACCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:68	14_7A1	ATGATTGAAGTCAAACCAATAAACGCGGAGGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCTAAAACAGTATCAGCTGAGAGGGATGGCGAC ACTCGAAGGGTACCGTGAGCAAAAAGCGGGAAGTACG CTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACGTCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGACCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:69	14_7A9	ATGATTGAAGTCAAACCAATAAACGCGGAGGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGTTGGTC AGCATCGCCTCCTTTCATCAAGCCAAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGGTACCGTGAGCAAAAAGCGGGTAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACGTCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:70	14_7G1	ATGATTGAAGTCAAACCAATAAACGCAGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGTTGAGAGGGATGGCGA CACTTGAAGAGTACCGTGAGCAAAAAGCGGGAAGTACG CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCGA

		GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:71	14_7H9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGATACCGTGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAA GGCGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGGTCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:72	14_8F7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCTGAAGCGCTTCTTCGGAAAAAAG GCGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCA AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACTGGGCCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:73	15_10C2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGAAGTACG CTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACAACTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGT GAAGTCTTCGACATACCGCCGACCGGACCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:74	15_10D6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTAGGTGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGGTCGGACCTCATATTT

		TGATGTATAAGAAATTGACGTAA
SEQ ID NO:75	15_11F9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTTGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAGAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:76	15_11H3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACACCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCAACTGGGCCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:77	15_12A8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:78	15_12D6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACACACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAGTTGACGTAA
SEQ ID	15_12D8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA

NO:79		TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAACTGAGAGGGATGGCG ACACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACGTCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CAAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:80	15_12D9	ATGATTGAAGTCAAACCAATAAACGCGGAGGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTCGAAGAGTACCGCGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGGTCGGACCTCATATTT TGATGTATAAGAAATTGACATAA
SEQ ID NO:81	15_3F10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTTGGTGGATATTACCGGGGCAAGCTGATC AGCATCGTTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGCACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACACACCGCCGGCCGGACCTCATATTTT GATGTATACGAAATTGACGTAA
SEQ ID NO:82	15_3G11	ATGATTGAAGTTAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TTGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTGTGGTGCAACGCCAGGACGTCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:83	15_4F11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TAAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC

		ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAGAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:84	15_4H3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGTA CGCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAA GGCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGACTGGGCCCCATATT TTGATGTATAAGAAATTGACGTAA
SEQ ID NO:85	15_6D3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACACCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:86	15_6G11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CAAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAGTTGACGTAA
SEQ ID NO:87	15_9F6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT

		TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTCGAAGAGTACCGCGAGCAAAAAGCGGGCAGTA CGCTTATCCGCCATGCCGAAGAGCTTCTTCGGAGAAAA GGCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCTGTCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:88	15F5	ATGATCGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGGTACTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCT TGAGGGCGAAGAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCTATGCCGAAGAGCTTCTTCGAAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGTG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:89	16A1	ATGATTGAAGTCAAACCTATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGCTTCACCTCGGTGGATATTACCAGGGCAAGCTGAT CAGCATCGCTTCCTTTCATAAAGCCGAACATTCAGGGCT TGAGGGCGAAGAACAGTATCAGCTGAGAGGGATGGCG ACGCTCGAAGGGTACCGCGAGCAAAAAGCGGGCAGTA CGCTTATCCGCCATGCCGAAGAGCTTCTTCGAAAAAAA GGCGCGGACCTTTTATGGTGCAATGCCAGGACATCTGT GAGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGATCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:90	16H3	ATGATTGACGTCAAACCTATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCAGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTTTTATGGTGCAATGCCAGGACATCTGT GAGCGGGTACTATGAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGATCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:91	17C12	ATGATTGAAGTCAAACCAATAAGCGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC

		GCTTATCCGCCATGCCGAAGAGCTTCTTCGAAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGTG AGCGGGTACTATGAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:92	18D6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCAA CGCTTGAAGGATACCGTGAGCAAAAAGCGGGAAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGGCATAA
SEQ ID NO:93	19C6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC TGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAATGCCAGGACATCTGTG AGAGGCTACTATGAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGATATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGGCGTAA
SEQ ID NO:94	19D5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACTGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAATGCCAGGACATCTGTGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:95	20A12	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCTTCCTTTCATAATGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGTGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGTAGACCTTTTATGGTGCAACGCCAGGACATCTGTG

		AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGATCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGGCATAA
SEQ ID NO:96	20F2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTG AGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:97	2.10E+12	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGATACCGTGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:98	23H11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAGGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTCCGAAAAAAAGG CGCGGACCTTTTATGGTGCAATGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCACCGATCGGACCTCATATTTTG ATGTATAAGAAATTGGCATAA
SEQ ID NO:99	24C1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTATCGGGACAGGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT

		GATGTATAAGAAACTGACGTAA
SEQ ID NO:100	24C6	ATGATTGAAGTCAAACCTATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGAAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGATATCTGTG AGCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGGCATAA
SEQ ID NO:101	2.40E+08	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAGGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCATCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATAATGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGATACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAATGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGGCATAA
SEQ ID NO:102	2_8C3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGTATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTATCGGGACAGGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:103	2H3	ATGATTGAAGTCAAACCGATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATC AGCACCGCTTCCTTTCATCAAGCCGGACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCGAAAAGCGGGAAGTAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGATATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGATATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID	30G8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA

NO:104		TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTTTGAAACCGATTTGCTCGGGGGTG CGTTTCACCTCGGTGGATATTACCAGGGCAAGCTGATCA GCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTTG AAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGAC GCTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACGC TTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGGC GCAGACCTTTTATGGTGCAACGCCAGGACATCTGTGAG CGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGCG AAGTCTACGACATACCGCCGATCGGACCTCATATTTTGA TGTATAAGAAATTGACGTAA
SEQ ID NO:105	3B_10C4	ATGATTGAAGTCAGACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGTATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGCCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:106	3B_10G7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGATCGGACCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:107	3B_12B1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:108	3B_12D10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGTATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTACGAAACCGATTTGCTCGGGGGT

		GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCCAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGATATCTGCGA GCGGGTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:109	3B_2E5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCAAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:110	3C_10H3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGTATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGAAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGATATCTGCGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:111	3C_12H10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGGGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:112	3C_9H8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGTATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTATCAGGACAGGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT

		GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCTATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGATATCTGCG AGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:113	4A_1B11	ATGATTGAAGTCAAACCTATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:114	4A_1C2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTATCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:115	4B_13E1	ATGATTGAAGTCAAACCTATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTGTGGTGCAACGCCAGGATATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:116	4B_13G10	TTACGTCAATTTCTTATACATCAAAATATGAGGTCCGAT CGGCGGTATGTCGTAGACTTCGCCCTGTTCGCTGAAGCC GAGCTTTTTATAGTACCCGCTCGCAGATGTCCTGGCGTT GCACCATAAAAGGTCCGCGCCTTTTTTCCGAAGAAGCTC TTCGGCATGGCGGATGAGCGTGCTTCCCGCTTTTTGCTC GCGGTACCCTTCAAGCGTCGCCATCCCTCTCAGCTGATA CTGTTTTTGGCCTTCAAGCTCTGAATGTTCGGCTTGATG

		AAAGGAGGCGATGCTGATCAGCTTGCCCCGGTAATATC CACCGAGGTGAAACGTGCCCCCGAGCAAATCAGTTTCA TACTTGCATGCTTCCAGCGGCTGATTCGGCCGGAGAATG CGGTGCCTGATCTCATACGTATCTTCCGCGTTTATTGGT TTGGCTTCAATCAT
SEQ ID NO:117	4B_16E1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGATATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:118	4B_17A1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATCAAGCCGAGCATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACATAA
SEQ ID NO:119	4B_18F11	ATGATTGAAGTCAATCCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTCTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATAATGCCGAACATTCAGAGCT TGATGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTG AGCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTC GATGTATAAGAAATTGACGTAA
SEQ ID NO:120	4B_19C8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGC

		GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAG GCGGGGTCTACGATATACCGCCGATCGGACCTCATATTT TGATGTATAAGAAATTGGCATAA
SEQ ID NO:121	4B_1G4	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAATCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGGTACCGCGAGCTAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGATATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:122	4B_21C6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGATATCTGCG AGCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGATATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:123	4B_2H7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTACCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGGCATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACATAA
SEQ ID NO:124	4B_2H8	ATGATTGAAGCCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACTGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT

		GATGTATAAGAAATTGACGTAA
SEQ ID NO:125	4B_6D8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGTAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACATGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:126	4B_7E8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:127	4C_8C9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTAACATAA
SEQ ID NO:128	4H1	ATGATTGAGGTGAAACCGATTAACGCAGAGGAGACCTA TGAACTAAGGCATAGGATACTCAGACCACACCAGCCGA TAGAGGTTTGTATGTATGAAACCGATTTACTTCGTGGTG CGTTTCACTTAGGCGGCTTTTACAGGGGCAAGCTGATTT CCATAGCTTCATTCCACCAGGCCGAGCATCCAGAACTCC AGGGCCAGAAACAATACCAACTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGACCAGAAAGCGGGATCGAGCCT AATTAAACACGCTGAACAGATCCTTCGGAAGCGGGGGG CGGACATGCTATGGTGCAATGCGCGGACATCCGCCGCT GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCGTAA TGTATAAACGCCTCACATAA
SEQ ID	6_14D10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA

NO:129		TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGAGGCAAGCTGATC AGCATCGCCTCCTTCCATCAAGCCGAACATTCAGAGCTT GAAGGCCATAAACAGTATCAGCTGAGAGGGATGGCGAC ACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCACG CTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:130	6_15G7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAA GGCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGGTCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:131	6_16A5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCACCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGAAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:132	6_16F5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGTACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:133	6_17C5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGCAAGTATGAAGCCGATTTGCTCGGGGGC

		ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAGCATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGAAACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACGTACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:134	6_18C7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAGGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTATCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGATATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTTTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:135	6_18D7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCMCCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:136	6_19A10	ATGATTGAAGCCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:137	6_19B6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTATCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT

		GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTCGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG CGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:138	6_19C3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:139	6_19C8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTACACCTCGGTGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCAAGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGGAATTGACGTAA
SEQ ID NO:140	6_20A7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCAGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATCAAGCCGAACATTCAGATCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGGTCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:141	6_20A9	ATGATTGAAGTCAAACCAATAAACGCGGGAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG

		CTTATCCGCCATGCCGAAGAGCTTCTACGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:142	6_20H5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGATACCGTGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:143	6_21F4	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCGTTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACGTACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:144	6_22C9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCGGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGGGCTT GAAGGCAAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACTTCCGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG AGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:145	6_22D9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGTATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCATGTATGAAACCGATTTGCTCGAGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAGCATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA

		GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:146	6_22H9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGATGAGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:147	6_23H3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGGAACTGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAGCAACCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAGCAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:148	6_23H7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGATACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCAGAAGAGATTCTTCGGAAAAAAG GCGCGGACCTCTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:149	6_2H1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCGTTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACCGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAATCTACGACATACCGCCGATCGGACCTCATATTTTG

		ATGTATAAGAAATTGACGTAA
SEQ ID NO:150	6_3D6	ATGATTGAAATCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGAGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CTCTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:151	6_3G3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:152	6_3H2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACATAA
SEQ ID NO:153	6_4A10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAACTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGATACCGTGAGCAAAAAGCGGGAAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID	6_4B1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA

NO:154		TGAGATCAGGCACCGCGTACTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC GGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGGACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACATAA
SEQ ID NO:155	6_5D11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:156	6_5F11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTAATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCCACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:157	6_5G9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTAAT CAGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGAGTACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGATATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:158	6_6D5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATGCGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACTGATTTGCTCGGGGGC

		ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:159	6_7D1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:160	6_8H3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:161	6_9G11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGCTTCACCTCGGTGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTA CGCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAA GGCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAG GCGAAGTCTACGACATACCGCCGGTCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:162	6F1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC TGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCTT

		GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGATGGATACCGCGAGCAAAAAGCGGGAAGCACG CTCATCCGCCATGCCGAAGAGCTTCTTCGAAAAAAAGG CGCGGACCTTTTATGGTGCAATGCCAGGACATCTGTGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:163	7_1C4	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAGCATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGATATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:164	7_2A10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACTGATTTGCTCGGGGGC ACGTTTCATCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGATATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:165	7_2A11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:166	7_2D7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGTGAGCAAAAAGCGGGAAGTACG

		CTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GTGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:167	7_5C7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGTGGGAAGCACG CTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGATATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:168	7_9C9	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAAATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTCATCCGCCATGCCGAAGAGCTTCTACGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:169	9_13F10	ATGATTGAAGTCAAACCAATAAACGCGGAGGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTTGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACTGGGCCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:170	9_13F1	ATGATTGAAGTCAAACCAATAAACGCGGAGGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTTGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCG

		AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACTGGGCCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:171	9_15D5	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGACGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTCTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:172	9_15D8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGGT CAGCATCGCCTCCTTTCATCAAGCTGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGCGCTTCTTCGGAAGAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACACACCGCCGGTCGGACCCCATATTTT GATGTATAAGAAGTTGACGTAA
SEQ ID NO:173	9_15H3	ATGATTGAAGTCAAGCCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATATGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCACGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTTAGCGAACAGGG CGAAGTCTACAACACACCGCCGGTTGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:174	9_18H2	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGTAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACA CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGGTCGGACCTCATATTTTG

		ATGTATAAGAAATTGACGTAA
SEQ ID NO:175	9_20F12	ATGATTGAAGTAAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCGTTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCGAGCTGGTC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCGGACCTTTTGTGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGACATACCGCCGGTCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:176	9_21C8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGTATGTATGAAACTGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTCGAAGGATACCGCGAGCAAAAAGCGGGCAGTA CGCTAATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTCTTATGGTGCAACGCCAGGACATCTGC GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGATCAGG GCGAAGTCTACGACATACCGCCGGTCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:177	9_22B1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATAAGGCACCGCATCCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGGTC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACTTACCGCCGACCGGACCCCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:178	9_23A10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGCTTCACCTCGGTGGATATTACCGGGGCAAGCTGGT CAGCATTGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAGGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGCGGGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAATGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID	9_24F6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA

NO:179		TGAGATCAGGCACCGCATTCTCAGGCCGAATCAGCCGC TAGAAGCATGCAAGTATGAAACCGATTTGCTCAGGGGT GCGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGCGCTTCTTCGGAAAAAAGG CGCGGACCTTTTGTGGTGCAACGCCAGGACGTCTGCGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:180	9_4H10	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACTGATTTGCTAGGGGGT ACGCTTCACCTCGGTGGATATTACCGGGGCAAGCTGAT CAGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GCGCGGACCTTATATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACATAA
SEQ ID NO:181	9_4H8	ATGATTGAAGTCAAACCAATAAATGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGAGGC ACGTTTCACCTAGGTGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTAATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACACTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACATAA
SEQ ID NO:182	9_8H1	ATGATTGAAGTCAAACCAATAACCGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTAGAAGGGTACCGCGAGCAAAAAGCGGGCAGTAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGAACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGACCGGACCCCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:183	9_9H7	ATGATTGAAGTCAAACCAATAAACGCGGAAGATGCGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGAGC

		ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGAGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCTGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:184	9C6	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC TGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAATGCCAGGACATCTGTG AGAGGCTACTATGAAAAGCTCGGCTTCAGCGAACAAGG CGGGGTCTACGATATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGGCGTAA
SEQ ID NO:185	9H11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGT TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGT ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATAAAGCCGAACATTCAGAGCT TGAGGGCGAAGAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGATACCGTGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAG GGGGCAGACCTTTTATGGTGCAATGCCAGGACATCTGT GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGATCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:186	0_4B10	ATGATAGAAGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCG AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGACTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACAACCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTTGATACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGCTCACATAA
SEQ ID NO:187	0_5B11	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCG

		AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGACTCT AATTAAACACGCTGAACAACTTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAAGATCACA
SEQ ID NO:188	0_5B3	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAACAACTTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACA
SEQ ID NO:189	0_5B4	ATGCTAGAGGTGAAACTGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGT TAGAAGCGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCG AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTTTCGTGATCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGAACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACA
SEQ ID NO:190	0_5B8	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGCTCACA
SEQ ID NO:191	0_5C4	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGCGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGGCCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTAT

		AATTAAACACGCTGAAGAAATTCTTCGTAAGAAGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACGTCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGACACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACA
SEQ ID NO:192	0_5D11	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGACTCT AATTAGACACGCTGAACAACTTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAGGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGCTCACA
SEQ ID NO:193	0_5D3	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGATCACATAA
SEQ ID NO:194	0_5D7	ATGATAGAAGTGAAACCGATTAACGCAGAGGAGACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTC GAAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTAC CTTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTC TAATTAGACACGCTGAACAACTTCTTCGTAAGAAGGGG GCGAATATGCTTTGGTGTAATGCGCGGACAACCGCCTC AGGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAG AGATATTTGATACGCCGCCAGTAGGACCTCACATCCTG ATGTATAAAAGGATCACA
SEQ ID NO:195	0_6B4	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CACTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTTTCGTGATCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG

		GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAAAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACA
SEQ ID NO:196	0_6D10	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGCTCACA
SEQ ID NO:197	0_6D11	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGCTTCACCTCGGTGGATATTACCGGGGCAAGCTGGT CAGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGGTACCGTGAGCAAAAAGCGGGCAGTAC GCTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGG GGGCAGACCTTTTATGGTGCAACGCCAGGACATCTGCG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGGTCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:198	0_6F2	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTTTCGTGAGCAGAAAGCGGGATCGACTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTTGATACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGATCACA
SEQ ID NO:199	0_6H9	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCG AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGACTCT AATTAGACACGCTGAAGAAATTCTTCGTAAGAAGGGGG CGAACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGACACGCCGCCAGTAGGACCTCACATCCTGATG

		TATAAAAGGCTCACA
SEQ ID NO:200	10_4C10	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTNTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACNTGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTTGATACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGCTCACATAA
SEQ ID NO:201	10_4D5	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGACTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACATAA
SEQ ID NO:202	10_4F2	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTTTGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGTAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGCTCACATAA
SEQ ID NO:203	10_4F9	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTTTCGTGAGCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGCTCACATAA
SEQ ID	10_4G5	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA

NO:204		TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTTTGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTACCGCGATCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGCTCACATAA
SEQ ID NO:205	10_4H4	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACATAA
SEQ ID NO:206	11_3A11	ATGATAGAAGTGAAACCGATTAACGCAGAGGATACCTA TGAACTGAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCGTCATTCCACCAGGCCGAGCACCCAGACCTC CAAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTAC CTTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTC TAATTAAACACGCTGAACAAATTCTTCGTAAGAGGGGG GCGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGCTCACATAA
SEQ ID NO:207	11_3B1	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTGAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTTTGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAACTCCGAGGTATGGCTACC TTGGAAGGTTTTCGTGAGCAGAAAGCGGGATCGACTCT AATTAGACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAGGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGACACGCCGCCAGTAGGGCCTCACATCCTGATG TATAAAAGGCTCACATAA
SEQ ID NO:208	11_3B5	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTTTGAAAGCGATTTACTTCGTGGTG

		CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCGTCATTCCACCAGGCCGAGCACTCGGAACTC CAAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTAC CTTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTC TAATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGG GCGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTC AGGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAG AGGTATTTGATACGCCGCCAGTAGGACCTCACATCCTG ATGTATAAAAGGATCACATAA
SEQ ID NO:209	11_3C12	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTGGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCGTCATTCCACCAGGCCGAGCACCCAGACCTC CAAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTAC CTTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTC TAATTAGACACGCTGAACAACTTCTTCGTAAGAGGGGG GCGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTCGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGATCACATAA
SEQ ID NO:210	11_3C3	ATGATAGAAGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CACTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCGTCATTCCACCAGGCCGAGCACTCAGAACTC CAAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTAC CTTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTC TAATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGG GCGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGACACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGATCACATAA
SEQ ID NO:211	11_3C6	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTTTGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCG AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGACTCT AATTAGACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACATAA
SEQ ID NO:212	11_3D6	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC

		AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGCTCACATAA
SEQ ID NO:213	1_1G12	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGCTCACATAA
SEQ ID NO:214	1_1H1	ATGATAGAAGTGAAACCTATTAACGCAGAGGAGACTTA CGAACTTCGACACAAGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCCAATTGATC TCGATTGCGAGTTTCCACAAAGCTGAACACTCAGAACT GCAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCG ACCCTCGAAGGATTCCGTGAGCAGAAGGCTGGCTCTTC GCTTATTAGGCACGCCGAGGAGATACTACGGAATAAAG GGGCAGATCTGCTTTGGTGTAATGCACGCACGACAGCC TCCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCACGGC GAAGTTTTCGAAACCCCGCCGGTTGGGCCGCACATTCTT ATGTACAAAAGAATCACT
SEQ ID NO:215	1_1H2	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGTT AGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGCT CGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATCT CGATTGCGAGTTTCCACCAAGCTGAACACTCAGAACTG GAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGA CCCTCGAAGGATTCCGTGAGCAGAAGGCTGGCTCTTCG CTTATTAGGCACGCCGAGGAGATACTACGGAAAAGAGG GGCAGATCTGCTTTGGTGTAATGCACGCACGACAGCCG CCGGTTACTATAAAAAGCTTGGTTTTAGTGAGCAGGGC GAAATTTTCGACACCCCGCCGGTTGGGCCGCACATTCTT ATGTACAAAAGAATCACT
SEQ ID NO:216	1_1H5	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAAATTCGACACAGGATCCTGCGCCCTAATCAGCCGT TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATC TCGATTGCGAGTTTCCACCAAGCTGAACACTCAGACCTG GAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGA CCCTCGAAGGATACCGTGATCAGAAGGCTGGCTCTTCG

		CTTATTAGGCACGCCGAGCAGATACTACGGAAAAGAGG GGCAGATCTGCTTTGGTGCAATGCACGCACGACAGCCG CCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCAGGGC GAAGTTTTCGACACCCCGCCGGTTGGGCCGCACATTCTT ATGTACAAAAAACTCACT
SEQ ID NO:217	1_2A12	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATC TCGATTGCGAGTTTCCACCAAGCTGAACAGTCAGAACT GGAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCG ACCCTCGAAGGATACCGTGATCAGAAGGCTGGCTCTAC GCTTATTAAGCACGCCGAGGAGATACTACGGAAAAAAG GGGCAGATCTGCTTTGGTGCAATGCACGCACGTCAGCC GCCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCAGGG CGAAATTTTCGACACCCCGCCGGTTGGGCCGCACATTCT TATGTACAAAAGACTCACT
SEQ ID NO:218	1_2B6	ATGATAGAAGTGAAACCTATTAACGCAGAGGAGACTTA CGAACTTCGACACAAGATCCTGCGCCCTAATCAGCCGTT AGAGGCATGCATGTATGAAACCGATCTGCTGCGGGGCT CGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATCT CGATTGCGAGTTTCCACCAAGCTGAACACTCAGAACTG GAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGA CCCTCGAAGGATTCCGTGATCAGAAGGCTGGCTCTTCGC TTATTAAGCACGCCGAGGAGATACTACGGAAAAGAGGG GCAGATCTGCTTTGGTGCAATGCACGCACGTCAGCCTCC GGTTACTATAAAAAGCTTGGTTTTAGTGAGCAGGGCGA AATTTTCGAAACCCCGCCGGTTGGGCCGCACATTCTTAT GTACAAAAGACTCACT
SEQ ID NO:219	1_2C4	ATGCTAGAAGTGAAACCTATTAACGCAGAGGAGACTTA CGAACTTCGACACAAGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAACCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCCAATTGATC TCGATTGCGAGTTTCCACCAAGCTGAACACTCAGACCTG CAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGAC CCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTACGC TTATTAAGCACGCCGAGGAGCTACTACGGAAAAAAGGG GCAGATCTGCTTTGGTGCAATGCACGCACGACAGCCGC CGGTTACTATAAAAAGCTTGGTTTTAGTGAGCAGGGCG AAGTTTTCGACACCCCGCCGGTTGGGCCGCACATTCTTA TGTACAAAAAAATCACT
SEQ ID NO:220	1_2D2	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAAGATCCTGCGCCCTAATCAGCCGTT AGAGGCATGCATGTATGAAAGCGATCTGCTGCGGAGCG CATTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATCT CGATTGCGAGTTTCCACAAAGCTGAACACTCAGAACTG CAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGAC CCTCGAAGGATACCGTGATCAGAAGGCTGGCTCTTCGC TTATTAGGCACGCCGAGGAGATACTACGGAAAAGAGGG GCAGATATGCTTTGGTGCAATGCACGCACGTCAGCCGC

		CGGTTACTATAAAAGGCTTGGTTTTAGTGAGCAGGGCG AAGTTTTCGACACCCCGCCGGTTGGGCCGCACATTCTTA TGTACAAAAGAATCACTTAA
SEQ ID NO:221	1_2D4	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATC TCGATTGCGAGTTTCCACCAAGCTGAACACTCAGACCTG CAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGAC CCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTTCGC TTATTAAGCACGCCGAGCAGCTACTACGGAAAAAAGGG GCAGATATGCTTTGGTGTAATGCACGCACGTCAGCCGC CGGTTACTATAAAAGGCTTGGTTTTAGTGAGCACGGCG AAATTTTCGAAACCCCGCCGGTTGGGCCGCACATTCTTA TGTACAAAAGAATCACT
SEQ ID NO:222	1_2F8	ATGCTAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGTT AGAGGCATGCATGTATGAAACCGATCTGCTGCGGGGCT CGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATCT CGATTGCGAGTTTCCACCAAGCTGAACATTCAGAACTG GAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGA CTCTCGAAGGATACCGTGATCAGAAGGCTGGCTCTTCG CTTATTAGGCACGCCGAGGAGATACTACGGAAAAGAGG GGCAGATATGCTTTGGTGCAATGCACGCACGACAGCCG CCGGTTACTATAAAAAGCTTGGTTTTAGTGAGCAGGGC GAAATTTACGACACCCCGCCGGTTGGGCCGCACATTCTT ATGTACAAAAAACTCACT
SEQ ID NO:223	1_2H8	ATGATAGAAGTGAAACCTATTAACGCAGAGGAGACTTA CGAACTTCGACACAAGATCCTGCGCCCTAATCAGCCGTT AGAGGCATGCATGTATGAAACCGATCTGCTGCGGGGCG CGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATCT CGATTGCGAGTTTCCACCAAGCTGACCACTCAGAACTG CAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGAC CCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTACGC TTATTAGGCACGCCGAGCAGATACTACGGAAAAGAGGG GCAGATCTACTTTGGTGCAATGCACGCACGTCAGCCGC CGGTTACTATAAAAAGCTTGGTTTTAGTGAGCACGGCG AAATTTTCGAAACCCCGCCGGTTGGGCCGCACATTCTTA TGTACAAAAGACTCACTTAA
SEQ ID NO:224	1_3A2	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC GCGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATC TCGATTGCGAGTTTCCACCAAGCTGAACACTCAGACCTG CAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGAC CCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTTCGC TTATTAGGCACGCCGAGGAGATACTACGGAAAAAAGGG GCAGATATGCTTTGGTGCAATGCACGCACGACAGCCGC CGGTTACTATAAAAGGCTTGGTTTTAGTGAGCAGGGCG AAGTTTTCGACACCCCGCCGGTTGGGCCGCACATTCTTA

		TGTACAAAAGAATCACT
SEQ ID NO:225	1_3D6	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGCTCACATAA
SEQ ID NO:226	1_3F3	ATGATAGAAGTGAAACCTATTAACGCAGAGGAGACTTA CGAACTTCGACAGAGGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCCAATTGATC TCGATTGCGAGTTTCCACCAAGCTGAACACTCAGAACT GCAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCG ACCCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTAC GCTTATTAAGCACGCCGAGGAGATACTACGGAAAAAAG GGGCAGATCTGCTTTGGTGCAATGCACGCACGTCAGCC GCCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCACGG CGAAATTTTCGACACCCCGCCGGTTGGGCCGCACATTCT TATGTACAAAAGAATCACT
SEQ ID NO:227	1_3H2	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAACCGATCTGCTGCGGGGC GCGTTCCATTTGGGCGGGTACTATCGTGGCCAATTGATC TCGATTGCGAGTTTCCACAAAGCTGAACACTCAGAACT GCAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCG ACCCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTAC GCTTATTAAGCACGCCGAGCAGCTACTACGGGAAAAAG GGGCAGATATGCTTTGGTGCAATGCACGCACGTCAGCC GCCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCAGGG CGAAGTTTTCGACACCCCGCCGGTTGGGCCGCACATTCT TATGTACAAAAAACTCACT
SEQ ID NO:228	1_4C5	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAAGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATC TCGATTGCGAGTTTCCACAAAGCTGAACACTCAGACCT GGAAGGGCAAAACCAGTATCAATTACGAGGGATGGCG ACCCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTAC GCTTATTAGGCACGCCGAGGAGATACTACGGAAAAGAG GGGCAGATATGCTTTGGTGCAATGCACGCACGTCAGCC TCCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCACGGC GAAATTTTCGACACCCCGCCGGTTGGGCCGCACATTCTT ATGTACAAAAGACTCACTTAA
SEQ ID	1_4D6	ATGCTAGAAGTGAAACCTATTAACGCAGAGGATACTTA

NO:229		CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAACCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCCAATTGATC TCGATTGCGAGTTTCCACAAAGCTGAACACTCAGACCT GGAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCG ACCCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTAC GCTTATTAGGCACGCCGAGCAGATACTACGGAAAAGAG GGGCAGATATGCTCTGGTGCAATGCACGCACGTCAGCC GCCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCAGGG CGAAGTTTTCGAAACCCCGCCGGTTGGGCCGCACATTCT TATGTACAAAAGACTCACT
SEQ ID NO:230	1_4H1	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGTT AGAGGCATGCATGTATGAAACCGATCTGCTGCGGGGCT CGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATCT CGATTGCGAGTTTCCACCAAGCTGAACACTCAGACCTG CAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGAC CCTCGAAGGATACCGTGAGCAGAAGGCTGGCTCTACGC TTATTAGGCACGCCGAGCAGCTACTACGGAAAAGAGGG GCAGATCTGCTTTGGTGCAATGCACGCACGTCAGCCTCC GGTTACTATAAAAGGCTTGGTTTTAGTGAGCACGGCGA AGTTTTCGACACCCCGCCGGTTGGGCCGCACATTCTTAT GTACAAAAGACTCACT
SEQ ID NO:231	1_5H5	ATGCTAGAAGTGAAACCTATTAACGCAGAGGAGACTTA CGAACTTCGACACAAGATCCTGCGCCCTAATCAGCCGTT AGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGCT CGTTCCATTTGGGCGGGTACTATCGTGGCCAATTGATCT CGATTGCGAGTTTCCACCAAGCTGAACACTCAGAACTG GAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGA CCCTCGAAGGATTCCGTGAGCAGAAGGCTGGCTCTACG CTTATTAAGCACGCCGAGCAGATACTACGGAAAAGAGG GGCAGATATGCTTTGGTGCAATGCACGCACGTCAGCCG CCGGTTACTATAAAAAGCTTGGTTTTAGTGAGCACGGC GAAATTTTCGACACCCCGCCGGTTGGGCCGCACATTCTT ATGTACAAAAAACTCACTTAA
SEQ ID NO:232	1_6F12	ATGATAGAAGTGAAACCTATTAACGCAGAGGAGACTTA CGAACTTCGACACAGGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC TCGTTCCATTTGGGCGGGTTCTATCGTGGCAAATTGATC TCGATTGCGAGTTTCCACCAAGCTGAACACTCAGACCTA GAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGA CCCTCGAAGGATACCGTGATCAGAAGGCTGGCTCTACG CTTATTAAGCACGCCGAGGAGCTACTACGGAAAAGAGG GGCAGATATGCTTTGGTGCAATGCACGCACGTCAGCCG CCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCACGGC GAAATTTACGAAACCCCGCCGGTTGGGCCGCACATTCTT ATGTACAAAAAAATCACT
SEQ ID NO:233	1_6H6	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACTTA CGAACTTCGACACAAGATCCTGCGCCCTAATCAGCCGA TAGAGGCATGCATGTATGAAAGCGATCTGCTGCGGGGC

		TCGTTCCATTTGGGCGGGTTCTATCGTGGCCAATTGATC TCGATTGCGAGTTTCCACCAAGCTGAACACTCAGACCTG GAAGGGCAAAAGCAGTATCAATTACGAGGGATGGCGA CCCTCGAAGGATACCGTGATCAGAAGGCTGGCTCTTCG CTTATTAAGCACGCCGAGGAGATACTACGGAAAAGAGG GGCAGATCTGCTTTGGTGCAATGCACGCACGTCAGCCG CCGGTTACTATAAAAGGCTTGGTTTTAGTGAGCAGGGC GAAATTTTCGACACCCCGCCGGTTGGGCCGCACATTCTT ATGTACAAAAAAATCACT
SEQ ID NO:234	3_11A10	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AGTTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGATCACATAA
SEQ ID NO:235	3_14F6	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACGTCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGCTCACATAA
SEQ ID NO:236	3_15B2	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGATCACATAA
SEQ ID NO:237	3_6A10	ATGATAGAAGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC

		AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGATCACATAA
SEQ ID NO:238	3_6B1	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACCCAGAACTC CAAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTAC CTTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTC TAATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGG GCGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGATCACATAA
SEQ ID NO:239	3_7F9	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGCGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGATCACATAA
SEQ ID NO:240	3_8G11	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCCAACCAGCCGA TAGAAGTGTGTATGTATGAAAGCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGATCACATAA
SEQ ID NO:241	4_1B10	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT

		AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGATCACATAA
SEQ ID NO:242	5_2B3	ATGATAGAAGTGAAACCTATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGTAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGATCACATAA
SEQ ID NO:243	5_2D9	ATGCTAGANGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGN TAGAAGTGTGTATGTATGAAANCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAACAAATTCTTCGTGAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGACACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGCTCACATAA
SEQ ID NO:244	5_2F10	ATGCTAGAAGTGAAACCTATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGCTCACATAA
SEQ ID NO:245	6_1A11	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCGTCATTCCACCAGGCCGAGCACTCAGACCTC CAAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTAC CTTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTC TAATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGG GCGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTC

		AGGCTACTACAGAAAGTTAGGCTTCAGCGAGCAGGGAG AGGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTG ATGTATAAAAGGCTCACATAA
SEQ ID NO:246	6_1D5	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGGGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGATCACATAA
SEQ ID NO:247	6_1F11	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGCTCACATAA
SEQ ID NO:248	6_1F1	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGCTCACATAA
SEQ ID NO:249	6_1H10	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCGGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGACACGCCGCCAGTAGGACCTCACATCCTGAT

		GTATAAAAAGATCACATAA
SEQ ID NO:250	6_1H4	ATGCTAGAAGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACGGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGATCAGAAAGCGGGATCGACTCT AATTAAACACGCTGAACAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GGTATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGCTCACATAA
SEQ ID NO:251	8_1F8	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACATAA
SEQ ID NO:252	8_1G2	ATGATAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAGAGTACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGCAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGAGACGCCGCCAGTAGGACCTCACATCCTGAT GTATAAAAGGCTCACGTAA
SEQ ID NO:253	8_1G3	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACTTA CGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG ATATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGATCACGTAA
SEQ ID	8_1H7	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA

NO:254		TGAACTAAGGCATAGAATACTCAGACCAAACCAGCCGA TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTTTTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGAACTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAAACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACATGCTTTGGTGCAATGCGCGGACATCCGCCTCA GGCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGA GATATTTGAAACGCCGCCAGTAGGACCTCACATCCTGA TGTATAAAAGGCTCACATAA
SEQ ID NO:255	8_1H9	ATGCTAGAGGTGAAACCGATTAACGCAGAGGATACCTA TGAACTAAGGCATAAAATACTCAGACCAAACCAGCCGT TAGAAGTGTGTATGTATGAAACCGATTTACTTCGTGGTG CATTTCACTTAGGCGGCTATTACAGGGGCAAACTGATTT CCATAGCTTCATTCCACCAGGCCGAGCACTCAGACCTCC AAGGCCAGAAACAGTACCAGCTCCGAGGTATGGCTACC TTGGAAGGTTATCGTGAGCAGAAAGCGGGATCGAGTCT AATTAGACACGCTGAAGAAATTCTTCGTAAGAGGGGGG CGGACTTGCTTTGGTGTAATGCGCGGACATCCGCCTCAG GCTACTACAAAAAGTTAGGCTTCAGCGAGCAGGGAGAG GTATTTGATACGCCGCCAGTAGGACCTCACATCCTGATG TATAAAAGGCTCACATAA
SEQ ID NO:256	GAT1_21F 12	ATGATTGAAGTCAAACCTATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGCGGATATTACCGGGGCAAGCTGAT CAGCATCGCTTCCTTTCATAATGCCGAACATTCAGAGCT TGAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCG ACGCTTGAAGGATACCGTGAGCAAAAAGCGGGAAGCA CGCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAA GGCGCGGACCTTTTATGGTGCAACGCCAGGACATCTGT GAGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGG GCGAAGTCTACGACATACCGCCGATCGGACCTCATATTT TGATGTATAAGAAATTGACGTAA
SEQ ID NO:257	GAT1_24G 3	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCCTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAATGCCAGGACATTTGTGA GCGGTTACTATGAAAAGCTCGGTTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTTATATTTTG ATGTATTAGAAATTGACATAA
SEQ ID NO:258	GAT1_29G 1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGT

		ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGTAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTGCGATATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGGCATAA
SEQ ID NO:259	GAT1_32G1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTACGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACATAA
SEQ ID NO:260	GAT2_15G8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TGGAAGCATGCAAGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATAATGCCGAACATTCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CGCTTGAAGGGTACCGCGAGCAAAAAGCGGGAAGCAC GCTCATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAG GCGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTG AGCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAGGG CGAAGTCTACGACATACCGCCGATCGGACCTCATATTTT GATGTATAAGAAATTGACGTAA
SEQ ID NO:261	GAT2_19H8	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATACTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATCCAGAGCTT GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGGTACCGCGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAAGG CGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTGA GCGGCTACTATGAAAAGCTCGGCTTCAGCGAACAGGGC GAAGTCTGCGACATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACATAA
SEQ ID NO:262	GAT2_21F1	ATGATTGAAGTCAAACCAATAAACGCGGAAGATACGTA TGAGATCAGGCACCGCATTCTCCGGCCGAATCAGCCGC TTGAAGCATGTATGTATGAAACCGATTTGCTCGGGGGC ACGTTTCACCTCGGTGGATATTACCGGGGCAAGCTGATC AGCATCGCTTCCTTTCATCAAGCCGAACATTCAGAGCTT

		GAAGGCCAAAAACAGTATCAGCTGAGAGGGATGGCGA CACTTGAAGGATACCGTGAGCAAAAAGCGGGCAGTACG CTTATCCGCCATGCCGAAGAGCTTCTTCGGAAAAAGGG GGCAGACCTTTTATGGTGCAACGCCAGGACATCTGTGA GCGGGTACTATAAAAAGCTCGGCTTCAGCGAACAAGGC GGGGTCTACGATATACCGCCGATCGGACCTCATATTTTG ATGTATAAGAAATTGACGTAA
SEQ ID NO:263	13_10F6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:264	13_12G6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQRQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:265	14_2A5	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGSTFHL GGYYRGKLISIASFNQAEHPELEGQKQYQLRGMATLEGYR EQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKKL GFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:266	14_2C1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPTGPHILMYKKLT
SEQ ID NO:267	14_2F11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPAGPHILMYKKLT
SEQ ID NO:268	CHIMERA	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRGAFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:269	10_12D7	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTLH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:270	10_15F4	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:271	10_17D1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:272	10_17F6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHSELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:273	10_18G9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHSELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK

		KLGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:274	10_1H3	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGRKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:275	10_20D10	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTLH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:276	10_23F2	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:277	10_2B8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:278	10_2C7	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:279	10_3G5	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:280	10_4H7	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:281	10_6D11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTLH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:282	10_8C6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGAFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:283	11C3	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYQGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:284	11G3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYQGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGGVYDIPPIGPHILMYKKLA
SEQ ID NO:285	11H3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGAFH LGGYYQGKLISIASFHKAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVRGYYEK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID	12_1F9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH

NO:286		LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVHDIPPTGPHILMYKKLT
SEQ ID NO:287	12_2G9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:288	12_3F1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:289	12_5C10	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDAPPTGPHILMYKKLT
SEQ ID NO:290	12_6A10	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:291	12_6D1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:292	12_6F9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:293	12_6H6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:294	12_7D6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPTGPHILMYKKLT
SEQ ID NO:295	12_7G11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:296	12F5	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYQGKLISIASFHKAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGGIYDIPPIGPHILMYKKLT
SEQ ID NO:297	12G7	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYQGKLISIASFHKAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:298	1_2H6	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGAFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK

		LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:299	13_12G12	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFNQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMHKKLT
SEQ ID NO:300	13_6D10	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDSLGGTFH LGGYYRGKLISIASFNQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:301	13_7A7	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRSAFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:302	13_7B12	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGSTFHL GGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGYR EQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKKL GFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:303	13_7C1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSARGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:304	13_8G6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDSLGGTFH LGGYYRGKLISIASFNQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:305	13_9F6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:306	14_10C9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:307	14_10H3	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:308	14_10H9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:309	14_11C2	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGSTFHL GGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPTGPHILMYKKLT
SEQ ID NO:310	14_12D8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYK KLGFREQGGVYDIPPVGPHILMYKKLT
SEQ ID	14_12H6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGAFH

NO:311		LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:312	14_2B6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFNQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:313	14_2G11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:314	14_3B2	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYK KLGFSEQGGVYDIPPAGPHILMYKKLT
SEQ ID NO:315	14_4H8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGSTFHL GGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGYR EQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKKL GFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:316	14_6A8	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLVSIASFNQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPVGPHVLMYKKLT
SEQ ID NO:317	14_6B10	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDMPPVGPHILMYKKLT
SEQ ID NO:318	14_6D4	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFNQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:319	14_7A11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGLKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPTGPHILMYKKLT
SEQ ID NO:320	14_7A1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPAGPHILMYKKLT
SEQ ID NO:321	14_7A9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAKHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:322	14_7G1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLISIASFNQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:323	14_7H9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK

		KLGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:324	14_8F7	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:325	15_10C2	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTTASGYYK KLGFSEQGEVFDIPPTGPHILMYKKLT
SEQ ID NO:326	15_10D6	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:327	15_11F9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFNQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRRKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:328	15_11H3	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:329	15_12A8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:330	15_12D6	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:331	15_12D8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGKVYDIPPVGPHILMYKKLT
SEQ ID NO:332	15_12D9	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:333	15_3F10	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLISIVSFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDTPPAGPHILMYTKLT
SEQ ID NO:334	15_3G11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:335	15_4F11	MIEVKPINAEDTYKIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLVSIASFNQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID	15_4H3	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH

NO:336		LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:337	15_6D3	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:338	15_6G11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGKVYDIPPVGPHILMYKKLT
SEQ ID NO:339	15_9F6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRRKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:340	15F5	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHKAEHSELEGEEQYQLRGMATLEGY REQKAGSTLIRYAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:341	16A1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTLH LGGYYQGKLISIASFHKAEHSGLEGEEQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:342	16H3	MIDVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYQGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:343	17C12	MIEVKPISAEDTYEIRHRILRPNQPLEACMYETDLLGGAFH LGGYYQGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:344	18D6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHKAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGEVYDIPPIGPHILMYKKLA
SEQ ID NO:345	19C6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLICIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVRGYYEK LGFSEQGGVYDIPPIGPHILMYKKLA
SEQ ID NO:346	19D5	MIEVKPINAEDTYEIRHCILRPNQPLEACMYETDLLGGTFH LGGYYQGKLISIASFHKAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:347	20A12	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYQGKLISIASFHNAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGVDLLWCNARTSVSGYYKK LGFSEQGGIYDIPPIGPHILMYKKLA
SEQ ID NO:348	20F2	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK

		LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:349	2.10E+12	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGAFH LGGYYQGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:350	23H11	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYQGKLISIASFHKAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGEVYDIPPIGPHILMYKKLA
SEQ ID NO:351	24C1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRDRLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:352	24C6	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARISVSGYYKKL GFSEQGGVYDIPPIGPHILMYKKLA
SEQ ID NO:353	2.40E+08	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHNAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGEVYDIPPIGPHILMYKKLA
SEQ ID NO:354	2_8C3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRDRLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:355	2H3	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYQGKLISTASFHQAGHSELEGQKQYQLRGMATLEG YRERKAGSTLIRHAEELLRKKGADLLWCNARISASGYYKK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:356	30G8	MIEVKPINAEDTYEIRHRILRPNQPLEACMFETDLLGGAFH LGGYYQGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:357	3B_10C4	MIEVRPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEAYDIPPIGPHILMYKKLT
SEQ ID NO:358	3B_10G7	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:359	3B_12B1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:360	3B_12D10	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGAFH LGGYYRGKLISIASFHPAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARISASGYYEKL GFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID	3B_2E5	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH

NO:361		LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSKQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:362	3C_10H3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARISASGYYKKL GFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:363	3C_12H10	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY RGQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:364	3C_9H8	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYQDRLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRYAEELLRKKGADLLWCNARISASGYYEKL GFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:365	4A_1B11	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:366	4A_1C2	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:367	4B_13E1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARISASGYYEKL GFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:368	4B_13G10	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPIGPYILMYKKLT
SEQ ID NO:369	4B_16E1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:370	4B_17A1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:371	4B_18F11	MIEVNPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTSH LGGYYRGKLISIASFHNAEHSELDGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVYDIPPIGPHISMYKKLT
SEQ ID NO:372	4B_19C8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPIGPHILMYKKLA
SEQ ID NO:373	4B_1G4	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGAFH LGGYYRGKLISIASFHQSEHPELEGQKQYQLRGMATLEGY RELKAGSTLIRHAEELLRKKGADLLWCNARISASGYYKKL

		GFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:374	4B_21C6	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARISASGYYKKL GFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:375	4B_2H7	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYGIPPIGPHILMYKKLT
SEQ ID NO:376	4B_2H8	MIEAKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:377	4B_6D8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEHGEVYDIPPIGPHILMYKKLT
SEQ ID NO:378	4B_7E8	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:379	4C_8C9	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRGAFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:380	4H1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGAFH LGGYYQGKLISIASFHQAVHSELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYK KLGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:381	6_14D10	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGHKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:382	6_15G7	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:383	6_16A5	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:384	6_16F5	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAVHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:385	6_17C5	MIEVKPINAEDTYEIRHRILRPNQPLEACKYEADLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGN REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDVPPIGPHILMYKKLT
SEQ ID	6_18C7	MIEVKPINAEDTYEIRHRILRPNQPLEACRYETDLLGGTFH

NO:386		LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARISASGYYKKL GFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:387	6_18D7	MIEVKPINAEDTYEIRXRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:388	6_19A10	MIEAKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:389	6_19B6	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRGAFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:390	6_19C3	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:391	6_19C8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTLH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRQAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKELT
SEQ ID NO:392	6_20A7	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRGTFH LGGYYRGKLISIASFHQAEHSDLEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:393	6_20A9	MIEVKPINAGDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:394	6_20H5	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:395	6_21F4	MIEVKPINAEDTYEIRHRVLRPNQPLEACMYETDLLGGAF HLGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDVPPVGPHILMYKKLT
SEQ ID NO:396	6_22C9	MIEVKPINAEDTYEIRHRILRPNRPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPGLEGKKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:397	6_22D9	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLEGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:398	6_22H9	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLDEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK

		LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:399	6_23H3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYGTDLLGGTFH LGGYYRGKLISIASFHQAEQPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:400	6_23H7	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEEILRKKGADLLWCNARTSASGYYKKL GFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:401	6_2H1	MIEVKPINAEDTYEIRHRVLRPNQPLEACMYETDLLGGTF HLGGYYRGKLISIASFHQAEHPELEGQKPYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEIYDIPPIGPHILMYKKLT
SEQ ID NO:402	6_3D6	MIEIKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFHL GGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGYR EQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKKL GFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:403	6_3G3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:404	6_3H2	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:405	6_4A10	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:406	6_4B1	MIEVKPINAEDTYEIRHRVLRPNQPLEACMYETDLLGGTF HLGGYYRGKLIGIASFHQAEHPELEGQKQYQLRGMATLE GYREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYY EKLGFSGQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:407	6_5D11	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:408	6_5F11	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVHDIPPVGPHILMYKKLT
SEQ ID NO:409	6_5G9	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARISASGYYKKL GFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:410	6_6D5	MIEVKPINAEDAYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID	6_7D1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLRGAFH

NO:411		LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:412	6_8H3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:413	6_9G11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTLH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:414	6F1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLVCIASFHKAEHSELEGQKQYQLRGMATLDG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYE KLGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:415	7_1C4	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:416	7_2A10	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:417	7_2A11	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:418	7_2D7	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:419	7_5C7	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKVGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:420	7_9C9	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:421	9_13F10	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLVSIASFHQAEHSELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:422	9_13F1	MIEAKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLVSIASFHQAEHTELEGQKQYQLRGMATLEE YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:423	9_15D5	MIEVKPINAEDTYEIRHRILRPNQPLDACKYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK

		LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:424	9_15D8	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDTPPVGPHILMYKKLT
SEQ ID NO:425	9_15H3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDMLRGAFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY HEQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYNTPPVGPHILMYKKLT
SEQ ID NO:426	9_18H2	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELVGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:427	9_20F12	MIEVKPINAEDTYEIRHRVLRPNQPLEACMYETDLLGGTF HLGGYYRGELVSIASFHQAEHPELEGQKQYQLRGMATLE GYREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYY KKLGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:428	9_21C8	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSDQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:429	9_22B1	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YREQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGEVYDLPPTGPHILMYKKLT
SEQ ID NO:430	9_23A10	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTLH LGGYYRGKLVSIASFHQAEHPELEGQKQYQLRGMATLEG YRGQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYK KLGFSEQGGVYDIPPVGPHILMYKKLT
SEQ ID NO:431	9_24F6	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLRGAFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEALLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:432	9_4H10	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTLH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLIWCNARTSASGYYKKL GFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:433	9_4H8	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFNQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID NO:434	9_8H1	MIEVKPITAEDTYEIRHRILRPNQPLEACKYETDLLGGTFHL GGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGYR EQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKKL GFSEQGEVYDIPPTGPHILMYKKLT
SEQ ID NO:435	9_9H7	MIEVKPINAEDAYEIRHRILRPNQPLEACKYETDLLGSTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEEY REQKAGSTLIRHAEELLRKKGADLLWCNARTSASGYYKK LGFSEQGEVYDIPPVGPHILMYKKLT
SEQ ID	9C6	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH

NO:436		LGGYYQGKLISIASFHNAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVYDIPPVGPHILMYKKLA
SEQ ID NO:437	9H11	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHKAEHSELEGEEQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:438	0_4B10	MIEVKPINAEDTYELRHKILRPNQPIEACMYESDLLRGAFH LGGFYRGKLISIASFHQAEHSDLEGQKQYQLRGMATLEGY RDQKAGSTLIKHAEEILRKRGADMLWCNARTTASGYYKK LGFSEQGEIFDTPPVGPHILMYKRLT
SEQ ID NO:439	0_5B11	MIEVKPINAEDTYELRHKILRPNQPIEACMYESDLLRGAFH LGGFYGGKLISIASFHQAEHSDLEGQKQYQLRGMATLEGY RDQKAGSTLIKHAEQLLRKRGADMLWCNARTSASGYYK KLGFSEQGEVFETPPVGPHILMYKKIT
SEQ ID NO:440	0_5B3	MLEVKPINAEDTYELRHRILRPNQPIEACMYETDLLRGAFH LGGFYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY RDQKAGSSLIKHAEQLLRKRGADLLWCNARTSASGYYKK LGFSEQGEVFDTPPVGPHILMYKRIT
SEQ ID NO:441	0_5B4	MLEVKLINAEDTYELRHRILRPNQPLEACMYETDLLRGAF HLGGFYRGKLISIASFHQAEHSDLEGQKQYQLRGMATLEG FRDQKAGSSLIKHAEEILRKRGANLLWCNARTSASGYYKK LGFSEQGEVFDTPPVGPHILMYKRIT
SEQ ID NO:442	0_5B8	MIEVKPINAEDTYELRHKILRPNQPIEACMYESDLLRGAFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY RDQKAGSSLIRHAEQILRKRGADLLWCNARTSASGYYKK LGFSEQGEIFDTPPVGPHILMYKRLT
SEQ ID NO:443	0_5C4	MIEVKPINAEDTYELRHKILRPNQPLEACMYETDLLRGAF HLGGFYRGKLISIASFHQAEHSGLQGQKQYQLRGMATLEG YREQKAGSSIIKHAEEILRKKGADLLWCNARTSASGYYKK LGFSEQGEIFDTPPVGPHILMYKRIT
SEQ ID NO:444	0_5D11	MIEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY REQKAGSTLIRHAEQLLRKRGADLLWCNARTSASGYYKR LGFSEQGEVFDTPPVGPHILMYKRLT
SEQ ID NO:445	0_5D3	MLEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY REQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:446	0_5D7	MIEVKPINAEETYELRHRILRPNQPIEACMYETDLLRGAFH LGGFYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY RDQKAGSSLIRHAEQLLRKKGANMLWCNARTTASGYYK KLGFSEQGEIFDTPPVGPHILMYKRIT
SEQ ID NO:447	0_6B4	MLEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGALH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGF RDQKAGSSLIRHAEQILRKRGADLLWCNARTSASGYYKK LGFSEQGKVFDTPPVGPHILMYKRIT
SEQ ID NO:448	0_6D10	MLEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YRDQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYK

		KLGFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:449	0_6D11	MIEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGYYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGF RDQKAGSSLIRHAEQILRKRGADLLWCNARTSASGYYKK LGFSEQGEVFETPPVGPHILMYKRIT
SEQ ID NO:450	0_6F2	MIEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGF REQKAGSTLIRHAEQILRKRGADMLWCNARTSASGYYKK LGFSEQGEIFDTPPVGPHILMYKRIT
SEQ ID NO:451	0_6H9	MIEVKPINAEDTYELRHKILRPNQPIEACMYETDLLRGAFH LGGFYGGKLISIASFHQAEHSDLEGQKQYQLRGMATLEGY REQKAGSTLIRHAEEILRKKGANLLWCNARTSASGYYKKL GFSEQGEVFDTPPVGPHILMYKRLT
SEQ ID NO:452	10_4C10	MIEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGXYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEG YRDQKAGSSLIKHAEQILRKRGADXLWCNARTSASGYYK KLGFSEQGEIFDTPPVGPHILMYKRLT
SEQ ID NO:453	10_4D5	MIEVKPINAEDTYELRHRILRPNQPIEVCMYETDLLRGAFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY REQKAGSTLIRHAEQILRKRGADLLWCNARTSASGYYKKL GFSEQGEVFDTPPVGPHILMYKRIT
SEQ ID NO:454	10_4F2	MLEVKPINAEDTYELRHRILRPNQPIEACMFESDLLRGAFH LGGFYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY REQKAGSSLIRHAEEILRKRGADMLWCNARTSASGYYKK LGFSEQGEIFETPPVGPHILMYKRLT
SEQ ID NO:455	10_4F9	MIEVKPINAEDTYELRHRILRPNQPIEVCMYETDLLRGAFH LGGFYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGF REQKAGSSLIRHAEQILRKRGADLLWCNARTSASGYYKKL GFSEQGEIFDTPPVGPHILMYKRLT
SEQ ID NO:456	10_4G5	MIEVKPINAEDTYELRHRILRPNQPIEACMFESDLLRGAFH LGGYYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YRDQKAGSSLIRHAEQILRKRGADLLWCNARTSASGYYK KLGFSEQGEIFDTPPVGPHILMYKRLT
SEQ ID NO:457	10_4H4	MLEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGFYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEG YREQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKK LGFSEQGEVFDTPPVGPHILMYKRIT
SEQ ID NO:458	11_3A11	MIEVKPINAEDTYELRHKILRPNQPIEVCMYESDLLRGAFH LGGFYRGKLISIASFHQAEHPDLQGQKQYQLRGMATLEGY RDQKAGSSLIKHAEQILRKRGADLLWCNARTSASGYYKK LGFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:459	11_3B1	MLEVKPINAEDTYELRHRILRPNQPIEACMFETDLLRGAFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGF REQKAGSTLIRHAEEILRKRGADLLWCNARTSASGYYKRL GFSEQGEIFDTPPVGPHILMYKRLT
SEQ ID NO:460	11_3B5	MIEVKPINAEDTYELRHRILRPNQPIEACMFESDLLRGAFH LGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY RDQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYKK LGFSEQGEVFDTPPVGPHILMYKRIT
SEQ ID	11_3C12	MIEVKPINAEDTYELRHRILRPNQPLEVCMYETDLLRGAFH

NO:461		LGGFYGGKLISIASFHQAEHPDLQGQKQYQLRGMATLEGY RDQKAGSSLIRHAEQLLRKRGADLLWCNARTSASGYYKK LGFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:462	11_3C3	MIEVKPINAEDTYELRHKILRPNQPIEACMYESDLLRGALH LGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY REQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEVFDTPPVGPHILMYKRIT
SEQ ID NO:463	11_3C6	MLEVKPINAEDTYELRHKILRPNQPIEACMFESDLLRGAFH LGGFYGGKLISIASFHQAEHSDLEGQKQYQLRGMATLEGY REQKAGSTLIRHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEIFDTPPVGPHILMYKRIT
SEQ ID NO:464	11_3D6	MIEVKPINAEDTYELRHRILRPNQPIEVCMYETDLLRGAFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY REQKAGSSLIKHAEQILRKRGADLLWCNARTSASGYYKKL GFSEQGEVFDTPPVGPHILMYKRLT
SEQ ID NO:465	1_1G12	MLEVKPINAEDTYELRHRILRPNQPIEVCMYETDLLRGAFH LGGFYGGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY RDQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:466	1_1H1	MIEVKPINAEETYELRHKILRPNQPIEACMYESDLLRGSFH LGGFYRGQLISIASFHKAEHSELQGQKQYQLRGMATLEGF REQKAGSSLIRHAEEILRNKGADLLWCNARTTASGYYKRL GFSEHGEVFETPPVGPHILMYKRIT
SEQ ID NO:467	1_1H2	MIEVKPINAEDTYELRHRILRPNQPLEACMYESDLLRGSFH LGGFYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGF REQKAGSSLIRHAEEILRKRGADLLWCNARTTAAGYYKK LGFSEQGEIFDTPPVGPHILMYKRIT
SEQ ID NO:468	1_1H5	MIEVKPINAEDTYEIRHRILRPNQPLEACMYESDLLRGSFH LGGFYRGKLISIASFHQAEHSDLEGQKQYQLRGMATLEGY RDQKAGSSLIRHAEQILRKRGADLLWCNARTTAAGYYKR LGFSEQGEVFDTPPVGPHILMYKKLT
SEQ ID NO:469	1_2A12	MIEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGSFH LGGFYRGKLISIASFHQAEQSELEGQKQYQLRGMATLEGY RDQKAGSTLIKHAEEILRKKGADLLWCNARTSAAGYYKR LGFSEQGEIFDTPPVGPHILMYKRLT
SEQ ID NO:470	1_2B6	MIEVKPINAEETYELRHKILRPNQPLEACMYETDLLRGSFH LGGFYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGF RDQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEIFETPPVGPHILMYKRLT
SEQ ID NO:471	1_2C4	MLEVKPINAEETYELRHKILRPNQPIEACMYETDLLRGSFH LGGFYRGQLISIASFHQAEHSDLQGQKQYQLRGMATLEGY REQKAGSTLIKHAEELLRKKGADLLWCNARTTAAGYYKK LGFSEQGEVFDTPPVGPHILMYKKIT
SEQ ID NO:472	1_2D2	MIEVKPINAEDTYELRHKILRPNQPLEACMYESDLLRSAFH LGGFYRGKLISIASFHKAEHSELQGQKQYQLRGMATLEGY RDQKAGSSLIRHAEEILRKRGADMLWCNARTSAAGYYKR LGFSEQGEVFDTPPVGPHILMYKRIT
SEQ ID NO:473	1_2D4	MIEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGSFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY REQKAGSSLIKHAEQLLRKKGADMLWCNARTSAAGYYK

		RLGFSEHGEIFETPPVGPHILMYKRIT
SEQ ID NO:474	1_2F8	MLEVKPINAEDTYELRHRILRPNQPLEACMYETDLLRGSF HLGGFYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEG YRDQKAGSSLIRHAEEILRKRGADMLWCNARTTAAGYYK KLGFSEQGEIYDTPPVGPHILMYKKLT
SEQ ID NO:475	1_2H8	MIEVKPINAEETYELRHKILRPNQPLEACMYETDLLRGAFH LGGFYRGKLISIASFHQADHSELQGQKQYQLRGMATLEGY REQKAGSTLIRHAEQILRKRGADLLWCNARTSAAGYYKK LGFSEHGEIFETPPVGPHILMYKRLT
SEQ ID NO:476	1_3A2	MIEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY REQKAGSSLIRHAEEILRKKGADMLWCNARTTAAGYYKR LGFSEQGEVFDTPPVGPHILMYKRIT
SEQ ID NO:477	1_3D6	MIEVKPINAEDTYELRHKILRPNQPIEACMYESDLLQGSFH LGGFYRGQLISIASFHQAEHSDLQGQKQYQLRGMATLEGF REQKAGSTLIKHAEEILRKKGADLLWCNARTSAAGYYKK LGFSEHGEIFDTPPAGPHILMYKKLT
SEQ ID NO:478	1_3F3	MIEVKPINAEETYELRQRILRPNQPIEACMYESDLLRGSFHL GGFYRGQLISIASFHQAEHSELQGQKQYQLRGMATLEGYR EQKAGSTLIKHAEEILRKKGADLLWCNARTSAAGYYKRL GFSEHGEIFDTPPVGPHILMYKRIT
SEQ ID NO:479	1_3H2	MIEVKPINAEDTYELRHRILRPNQPIEACMYETDLLRGAFH LGGYYRGQLISIASFHKAEHSELQGQKQYQLRGMATLEGY REQKAGSTLIKHAEQLLREKGADMLWCNARTSAAGYYK RLGFSEQGEVFDTPPVGPHILMYKKLT
SEQ ID NO:480	1_4C5	MIEVKPINAEDTYELRHKILRPNQPIEACMYESDLLRGSFH LGGFYRGKLISIASFHKAEHSDLEGQNQYQLRGMATLEGY REQKAGSTLIRHAEEILRKRGADMLWCNARTSASGYYKR LGFSEHGEIFDTPPVGPHILMYKRLT
SEQ ID NO:481	1_4D6	MLEVKPINAEDTYELRHRILRPNQPIEACMYETDLLRGSFH LGGFYRGQLISIASFHKAEHSDLEGQKQYQLRGMATLEGY REQKAGSTLIRHAEQILRKRGADMLWCNARTSAAGYYKR LGFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:482	1_4H1	MIEVKPINAEDTYELRHRILRPNQPLEACMYETDLLRGSFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY REQKAGSTLIRHAEQLLRKRGADLLWCNARTSASGYYKR LGFSEHGEVFDTPPVGPHILMYKRLT
SEQ ID NO:483	1_5H5	MLEVKPINAEETYELRHKILRPNQPLEACMYESDLLRGSFH LGGYYRGQLISIASFHQAEHSELEGQKQYQLRGMATLEGF REQKAGSTLIKHAEQILRKRGADMLWCNARTSAAGYYKK LGFSEHGEIFDTPPVGPHILMYKKLT
SEQ ID NO:484	1_6F12	MIEVKPINAEETYELRHRILRPNQPIEACMYESDLLRGSFHL GGFYRGKLISIASFHQAEHSDLEGQKQYQLRGMATLEGYR DQKAGSTLIKHAEELLRKRGADMLWCNARTSAAGYYKR LGFSEHGEIYETPPVGPHILMYKKIT
SEQ ID NO:485	1_6H6	MIEVKPINAEDTYELRHKILRPNQPIEACMYESDLLRGSFH LGGFYRGQLISIASFHQAEHSDLEGQKQYQLRGMATLEGY RDQKAGSSLIKHAEEILRKRGADLLWCNARTSAAGYYKR LGFSEQGEIFDTPPVGPHILMYKKIT
SEQ ID	3_11A10	MLEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH

NO:486		LGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY REQKAGSSLVKHAEEILRKRGADLLWCNARTSASGYYKK LGFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:487	3_14F6	MLEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGFYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY REQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEIFETPPVGPHILMYKRLT
SEQ ID NO:488	3_15B2	MLEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGYYGGKLISIASFHQAEHSELQGQKQYQLRGMATLE GYREQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYK KLGFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:489	3_6A10	MIEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY REQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:490	3_6B1	MLEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGYYRGKLISIASFHQAEHPELQGQKQYQLRGMATLEGY REQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEVFETPPVGPHILMYKRIT
SEQ ID NO:491	3_7F9	MLEVKPINAEDTYELRHRILRPNQPIEACMYESDLLRGAFH LGGYYGGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YREQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKK LGFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:492	3_8G11	MLEVKPINAEDTYELRHRILRPNQPIEVCMYESDLLRGAFH LGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY REQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:493	4_1B10	MIEVKPINAEDTYELRHRILRPNQPIEVCMYETDLLRGAFH LGGFYGGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY RDQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYKK LGFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:494	5_2B3	MIEVKPINAEDTYELRHRILRPNQPLEVCMYETDLLRGAFH LGGFYGGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY RDQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYKK LGFSEQGEIFETPPVGPHILMYKRIT
SEQ ID NO:495	5_2D9	MLXVKPINAEDTYELRHKILRPNQPXEVCMYEXDLLRGAF HLGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YRDQKAGSSLIKHAEQILRERGADMLWCNARTSASGYYK KLGFSEQGEVFDTPPVGPHILMYKRLT
SEQ ID NO:496	5_2F10	MLEVKPINAEDTYELRHKILRPNQPIEVCMYETDLLRGAF HLGGFYGGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YRDQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYK KLGFSEQGEIFETPPVGPHILMYKRLT
SEQ ID NO:497	6_1A11	MLEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YRDQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYR KLGFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:498	6_1D5	MIEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YRDQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYK

		KLGFSEQGEVFETPPVGPHILMYKRIT
SEQ ID NO:499	6_1F11	MIEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YREQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYK KLGFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:500	6_1F1	MLEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGFYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEG YRDQKAGSSLIRHAEQILRKRGADMLWCNARTSASGYYK KLGFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:501	6_1H10	MLEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGFYGGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YRDQKAGSSLIRHAEEILRKRGADMLWCNARTSASGYYK KLGFSEQGEVFDTPPVGPHILMYKKIT
SEQ ID NO:502	6_1H4	MLEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGFYGGKLISIASFHQAEHSDLQGQKQYQLRGMATLEG YRDQKAGSTLIKHAEQILRKRGADMLWCNARTSASGYYK KLGFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:503	8_1F8	MIEVKPINAEDTYELRHRILRPNQPLEVCMYETDLLRGAFH LGGFYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLEGY REQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKKL GFSEQGEIFDTPPVGPHILMYKRIT
SEQ ID NO:504	8_1G2	MIEVKPINAEDTYELRHRVLRPNQPLEVCMYETDLLRGAF HLGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEG YREQKAGSSLIKHAEEILRKRGADLLWCNARTSASGYYKK LGFSEQGEVFETPPVGPHILMYKRLT
SEQ ID NO:505	8_1G3	MLEVKPINAEDTYELRHKILRPNQPIEVCMYETDLLRGAF HLGGYYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEG YREQKAGSSLIRHAEEILRKRGADLLWCNARTSASGYYKK LGFSEQGEIFDTPPVGPHILMYKRIT
SEQ ID NO:506	8_1H7	MLEVKPINAEDTYELRHRILRPNQPIEVCMYETDLLRGAFH LGGFYRGKLISIASFHQAEHSELQGQKQYQLRGMATLEGY REQKAGSSLIKHAEEILRKRGADMLWCNARTSASGYYKK LGFSEQGEIFETPPVGPHILMYKRLT
SEQ ID NO:507	8_1H9	MLEVKPINAEDTYELRHKILRPNQPLEVCMYETDLLRGAF HLGGYYRGKLISIASFHQAEHSDLQGQKQYQLRGMATLE GYREQKAGSSLIRHAEEILRKRGADLLWCNARTSASGYYK KLGFSEQGEVFDTPPVGPHILMYKRLT
SEQ ID NO:508	GAT1_21F12	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHNAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:509	GAT1_24G3	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTFVSGYYEK LGFSEQGEVYDIPPIGPYILMYEKLT
SEQ ID NO:510	GAT1_29G1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGGVCDIPPIGPHILMYKKLA
SEQ ID	GAT1_32G	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH

NO:511	1	LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:512	GAT2_15G8	MIEVKPINAEDTYEIRHRILRPNQPLEACKYETDLLGGTFH LGGYYRGKLISIASFHNAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGEVYDIPPIGPHILMYKKLT
SEQ ID NO:513	GAT2_19H8	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHPELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYEK LGFSEQGEVCDIPPIGPHILMYKKLT
SEQ ID NO:514	GAT2_21F1	MIEVKPINAEDTYEIRHRILRPNQPLEACMYETDLLGGTFH LGGYYRGKLISIASFHQAEHSELEGQKQYQLRGMATLEGY REQKAGSTLIRHAEELLRKKGADLLWCNARTSVSGYYKK LGFSEQGGVYDIPPIGPHILMYKKLT
SEQ ID NO:515	*B. licheniformis* ribosome binding site	AACTGAAGGAGGAATCTC